

10/510 952

10 Rec'd PCT 29 OCT 2004

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international(43) Date de la publication internationale
13 novembre 2003 (13.11.2003)

PCT

(10) Numéro de publication internationale
WO 03/093461 A2(51) Classification internationale des brevets⁷ : C12N 7/00(21) Numéro de la demande internationale :
PCT/FR03/01328

(22) Date de dépôt international : 28 avril 2003 (28.04.2003)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
02/05424 30 avril 2002 (30.04.2002) FR

(71) Déposants (pour tous les États désignés sauf US) : INSTITUT PASTEUR [FR/FR]; 25-28, rue du Docteur-Roux, F-75015 Paris (FR). CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) [FR/FR]; 3, rue Michel Ange, F-75016 Paris (FR).

(71) Déposant (pour US seulement) : GENOSCOPE -CENTRE NATIONAL DE SEQUENCAGE [FR/FR]; 2, rue Gaston Crémieux, F-91000 Evry (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : MAR-LIERE, Philippe [FR/FR]; 2, allée Saint-Martin, F-91450 Etolles (FR). KAMINSKI, Pierre-Alexandre [FR/FR]; 4, rue Bailly, F-75003 Paris (FR). GALISSON, Frédérique [FR/CH]; 35, chemin de Bonne Espérance, CH-1006 Lausanne (CH). BOUZON, Madeleine [FR/FR]; 4, rue des Capucins, F-92190 Meudon (FR). POCHET, Sylvie [FR/FR]; 201, rue Lecourbe, F-75015 Paris (FR). WEISSENBAACH, Jean [FR/FR]; 163, rue de Vaugirard,

F-75015 Paris (FR). SAURIN, William [FR/FR]; 31, rue de la Procession, F-75015 Paris (FR). ROBERT, Catherine [FR/FR]; 2, rue Gaston Crémieux, F-91057 Evry (FR). VICO, Virginie [FR/FR]; 2, rue Gaston Crémieux, F-91057 Evry (FR).

(74) Mandataire : SANTARELLI; 14, avenue de la Grande-Armée, Boîte postale 237, F-75822 Paris Cedex 17 (FR).

(81) États désignés (national) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) États désignés (régional) : brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Publiée :

— sans rapport de recherche internationale, sera republiée dès réception de ce rapport

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

(54) Title: GENOMIC LIBRARY OF CYANOPHAGE S-2L AND FUNCTIONAL ANALYSIS

(54) Titre : BANQUE GENOMIQUE DU CYANOPHAGE S-2L ET ANALYSE FONCTIONNELLE

(57) Abstract: The invention relates to the genome sequence and the nucleotide sequences coding for polypeptides of cyanophage S-2L. According to the invention, the polypeptides include, but are not limited to, polypeptides involved in the synthesis, transcription and replication of purine bases. In particular, determining the genome of cyanophage S-2L is useful for supplying genes which, expressed in recombinant bacteria, enable the synthesis of DNA monomers incorporating base D (2,6 diaminopurine) instead of base A (adenine), thereby producing chemically-remodelled nucleic acids in the bacteria.

(57) Abrégé: L présente invention a pour objet la séquence génomique et des séquences nucléotidiques codant pour des polypeptides du cyanophage S-2L. Les polypeptides décrits dans la présente invention sont, de façon non limitative, des polypeptides impliqués dans la synthèse, la transcription et la réplication des bases puriques. En particulier, la détermination du génome du cyanophage S-2L est un outil utile pour la fourniture de gènes, qui exprimés dans des bactéries recombinantes, permettent la sythèse de monomères d'ADN incorporant la base D (2,6 diaminopurine) au lieu de la base A (adénine) et ainsi de produire des acides nucléiques chimiquement remodelés chez les bactéries.

WO 03/093461 A2

Banque génomique du cyanophage S-2L et analyse fonctionnelle

5 La présente invention a pour objet la séquence génomique et des séquences nucléotidiques codant pour des polypeptides du cyanophage S-2L. Les polypeptides décrits dans la présente invention sont, de façon non limitative, des polypeptides impliqués dans la synthèse, la transcription et la réplication des bases puriques. En particulier, la détermination du génome du cyanophage S-2L est un outil utile pour la
10 fourniture de gènes, qui exprimés dans des bactéries recombinantes, permettent la synthèse de monomères d'ADN incorporant la base D (2,6 diaminopurine) au lieu de la base A (adénine) et ainsi de produire des acides nucléiques chimiquement remodelés chez les bactéries.

L'invention concerne également l'utilisation de la séquence génomique et/ou
15 des séquences nucléotidiques et/ou polypeptidiques décrites dans la présente invention pour l'analyse de l'expression de gènes.

Les deux acides nucléiques principaux ADN et ARN sont des polymères de nucléotides qui sont composés d'une base purine ou pyrimidine liée à un sucre à 5 carbones (désoxyribose dans le cas de l'ADN, ribose dans l'ARN) par une liaison N-
20 glycosidique et un phosphate estérifié au groupement hydroxyl du carbone situé à la position 5' du sucre: ARN et ADN contiennent quatre types de nucléotides qui se distinguent par leurs bases : adénine (A), guanine (G), cytosine (C) et uracile (U) pour l'ARN ; le 5-méthyluracile, c'est-à-dire la thymine (T), remplaçant l'uracile dans l'ADN.

25 Parmi les altérations chimiques de l'ADN et de l'ARN possibles seules des modifications des bases et non du sucre ont été observées. Contrairement à l'ADN aucun ARN modifié n'a pu être répliqué à ce jour.

Des bases modifiées s'observent dans l'ADN de tous les organismes, et peuvent être impliquées dans des phénomènes de régulation de l'expression génétique (5).
30 Sauf chez les bactériophages, les modifications de l'ADN connues jusqu'à présent sont produites par des réactions enzymatiques post-répliquatives, dont un duplex d'ADN est le substrat.

Au contraire, lors de l'infection par certains bactériophages, les modifications de l'ADN connues jusqu'à présent sont produites par des réactions enzymatiques pré-réplicatives, dont un nucléotide est le substrat, pour aboutir à un désoxynucléoside triphosphate non-canonique. Parmi les modifications connues on citera les entités :
5 dUTP, 5-hydroxyméthyl-dUTP, 5-dihydroxypentyl-dUTP, 5-hydroxyméthyl-dCTP. Une autre entité est fortement suspectée : le 5-méthyl-dCTP (11). L'émergence des bases modifiées chez les bactériophages est généralement interprétée comme une contre-mesure aux systèmes de restriction des bactéries (11).

Quelques exemples de bases modifiées sont représentés sur la figure 1.

10 Le bromouracile ou la 8-azaguanine sont des analogues synthétiques des bases naturelles thymine et guanine. Ces analogues sont convertis en nucléotides triphosphate par les voies de sauvegarde des purines ou pyrimidines et sont ensuite incorporés dans l'ADN.

La 6-méthyladénine et la 5-méthylcytosine sont les bases modifiées les plus
15 fréquemment rencontrées. Les nucléotides méthylés ne sont pas incorporés en tant que tel dans l'ADN mais sont le produit de l'action d'ADN méthyltransférases spécifiques. Ces enzymes transfèrent le groupement méthyl de la S- adénosylméthionine à l'adénine ou la cytosine et ce après la réplication de l'ADN. Chez les procaryotes le rôle principal de la méthylation de l'ADN est la dégradation
20 de l'ADN étranger. Chez les eucaryotes la méthylation de l'ADN influe sur la régulation de l'expression des gènes et sur la différenciation cellulaire.

Chez certaines phages de type T comme le bactériophage T4, la cytosine est remplacée systématiquement par la 5-hydroxyméthylcytosine. Cette substitution nécessite d'une part une voie de biosynthèse de l'hydroxyméthyl-désoxycytidine tri-
25 phosphate (HM dCTP) ainsi que des enzymes permettant l'exclusion de la base normale.

La voie de biosynthèse de l'HMC DNA fait intervenir une hydroxyméthylase qui convertit le dCMP en hydroxyméthyl dCMP, une nucléoside monophosphate kinase qui phosphoryle le HM dCMP pour donner le diphosphate, précurseur du HM
30 dCTP qui est ensuite incorporé dans l'ADN polymérase puis glycosylé par une glycosyltransférase.

L'exclusion de la cytosine fait intervenir d'une part des endonucléases

spécifiques de l'ADN contenant cette base et une dCDPase-dCTPase qui convertit les nucléotides correspondant en dCMP qui est alors substrat de la dCMP hydroxyméthylase et de la dCMP désaminase. La dCMP désaminase génère le dUMP précurseur du dTMP.

- 5 Par un mécanisme similaire à celui décrit ci-dessus, la thymine est remplacée par le 5-hydroxyméthyluracile (phages SPO1 et ϕ e) ou l'uracile (phages PBS2) chez plusieurs phages de *Bacillus subtilis* (Warren, 1980 ; Kornberg et Baker, 1991).

D'autres phages tel SP15 ou ϕ W14 ont un ADN dont la thymine a été remplacée par la 5-dihydroxypentyluracile et l' α -putrescinylothymine. Cependant, ce
10 remplacement n'est que partiel et semble être dû à des modifications post-répliquatives.

Dans le cas de S-2L, la voie de synthèse de la base D n'est pas encore parfaitement établie et la modification post-répliquative de l'adénine en diaminopurine ne peut être totalement écartée. Cependant, la biosynthèse du monomère non-
15 canonique dDTP apparaît comme significativement plus vraisemblable, compte tenu du fait que le remplacement de A par D dans l'ADN de S-2L est total et non majoritaire (7,8), comme c'est le cas pour les modifications post-répliquatives d'hydroxyméthyl-U en putrescinyl-T dans le phage ϕ W14. De plus, la modification de A en D in situ nécessiterait la rupture des liaisons hydrogène des duplex d'ADN
20 et, pouvant difficilement s'effectuer en une seule étape chimique, introduirait des lésions mutagènes si ce processus était interrompu.

Le cyanophage S-2L

Le cyanophage S-2L a été isolé d'échantillons d'eau prélevés dans la région de
25 Léninegrad. Ce phage est capable de lyser un nombre relativement restreint de *Synechococcus*: sp. 698, 58 et PCC6907. D'un point de vue morphologique il se compose d'une tête icosaédrique et d'une queue flexible non contractile. S-2L ferait partie d'une famille dont l'autre membre pourrait être le phage SM-2 qui lui est morphologiquement similaire (Fox et coll. 1976).

30 L'ADN du phage S-2L est linéaire double brin d'une taille de 42 kb composé de 70% G:C et de 30% d'une paire équivalente à A:T dans laquelle l'adénine a été remplacée par la 2,6 diaminopurine (D). Ce remplacement est total et aucune autre

base n'a pu être identifiée (Kirmos et coll., 1977 ; Khudyokov et coll., 1978). Comme on l'a vu précédemment, seuls des remplacements totaux de bases pyrimidines ont été rapportés, S-2L est jusqu'à ce jour le seul cas pour une base purine.

Comme dans les paires G:C trois liaisons hydrogène sont formées entre la purine et la pyrimidine de la paire D:T ce qui confère à l'ADN une plus grande stabilité.

La présence de la base D dans l'ADN de S-2L entraîne une résistance à la digestion par les endonucléases de restriction possédant un A dans leur site de reconnaissance (l'enzyme de restriction TaqI étant la seule exception). Par contre, la paire D:T semble être reconnue comme une paire G:C par les enzymes de restriction clivant les séquences riches en G:C comme SmaI (Szekeres et Matveyev, A.V., 1978).

Compte tenu du fait que le remplacement de la base A par D est total et non majoritaire, il est fort vraisemblable que le génome du phage S-2L code pour au moins une voie de biosynthèse de la base D.

Au vu de l'art antérieur, l'étude du cyanophage S-2L demande de nouvelles approches, en particulier génétiques, afin d'améliorer la compréhension des différentes voies métaboliques de cet organisme.

Ainsi, un objet de la présente invention est de divulguer la séquence complète du génome du cyanophage S-2L et de tous les gènes contenus dans cedit génome.

En effet, la connaissance du génome de cet organisme permet de mieux définir les interactions entre les différents gènes, les différentes protéines, et par là-même, les différentes voies métaboliques. En effet, et contrairement à la divulgation de séquences isolées, la séquence génomique complète d'un organisme forme un tout, permettant d'obtenir immédiatement toutes les informations nécessaires à cet organisme pour croître et fonctionner.

L'invention vise en particulier à séquencer le génome du phage S-2L, de manière à obtenir un gisement de gènes qui, une fois propagés isolément et exprimés sous contrôle dans des bactéries recombinantes, sont destinés notamment à former par voie biotechnologique de nouveaux monomères de l'ADN et à produire, voir répliquer, des acides nucléiques chimiquement remodelés chez les bactéries.

L'invention vise également à utiliser des séquences nucléotidiques obtenues

pour l'identification des voies métaboliques conduisant à la production des bases D.

L'invention vise également à la production enzymatique d'analogues de désoxynucléosides très utiles notamment dans la chimiothérapie du SIDA.

L'invention vise également à exprimer chez un hôte du cyanophage S2L des
5 acides nucléiques codant pour des protéines impliquées dans le métabolisme des bases D.

Ainsi l'invention vise également à obtenir des gènes de S-2L qui propagés individuellement chez *E.coli* et exprimés sous strict contrôle transcriptionnel permettront de tester les hypothèses concernant leur fonction dans le métabolisme
10 des nucléotides, la réplication et la transcription.

Pour atteindre les différents résultats techniques recherchés, l'invention concerne selon un premier aspect une séquence nucléotidique du cyanophage S-2L correspondant à SEQ ID N° 1.

La présente invention concerne également une séquence nucléotidique du
15 cyanophage S-2L choisie parmi :

- a) une séquence nucléotidique comportant au moins 80 %, 85 %, 90 %, 95 % ou 98 % d'identité avec SEQ ID N° 1 ;
- b) une séquence nucléotidique hybridant dans des conditions de forte stringence avec SEQ ID N° 1 ;
- 20 c) une séquence nucléotidique complémentaire de SEQ ID N° 1 ou complémentaire d'une séquence nucléotidique telle que définie en a), ou b), ou une séquence nucléotidique de l'ARN correspondant ;
- d) une séquence nucléotidique de fragment représentatif de SEQ ID N° 1, ou de fragment représentatif d'une séquence nucléotidique telle que
25 définie en a), b) ou c);
- e) une séquence nucléotidique comprenant une séquence telle que définie en a), b), c) ou d) ; et
- f) une séquence nucléotidique modifiée d'une séquence nucléotidique telle que définie en a), b), c), d) ou e).

30 De façon plus particulière, la présente invention a également pour objet les séquences nucléotidiques caractérisées en ce qu'elles sont issues de SEQ ID N° 1 et en ce qu'elles codent pour des polypeptides choisis parmi les séquences SEQ ID N° 2

à SEQ ID N° 527 ou un fragment biologiquement actif de ces polypeptides.

De plus, l'invention concerne également les séquences nucléotidiques caractérisées en ce qu'elles comprennent une séquence nucléotidique choisie parmi :

- 5 a) une séquence nucléotidique issue de SEQ ID N° 1 et codant pour un polypeptide choisi parmi les séquences parmi SEQ ID N° 2 à SEQ ID N° 527.
- b) une séquence nucléotidique comportant au moins 80 %, 85 %, 90 %, 95 % ou 98 % d'identité avec une séquence nucléotidique selon a) ;
- 10 c) une séquence nucléotidique s'hybridant dans des conditions de forte stringence avec une séquence nucléotidique selon a) ou b);
- d) une séquence nucléotidique complémentaire ou d'ARN correspondant à une séquence telle que définie en a), b) ou c) ;
- e) une séquence nucléotidique de fragment représentatif d'une séquence telle que définie en a), b), c) ou d) ; et
- 15 f) une séquence nucléotidique modifiée d'une séquence telle que définie en a), b), c), d) ou e),

De préférence l'invention concerne une séquence nucléotidique caractérisée en ce qu'elle code pour un polypeptide choisi parmi :

- 20 a) les polypeptides du cyanophage S-2L de séquences SEQ ID N°2 à SEQ ID N°527 ;
- b) de préférence les 54 polypeptides mentionnés sur le tableau 1 à savoir: SEQ ID N°14,18,26,68,86,92,105,109,134,142,143,148,152,169,175,187, 208,211, 234,246, 250,257,264,286,298,316,332,342,347,348,351,355,364,365, 369,370,392,395, 406, 25 418,422,425,429,432,433,454,464,466,472,484,489,494,500 ;
- c) de préférence encore les 14 polypeptides du cyanophage S-2L indiqués sur le tableau 1 comme présentant une homologie significative à savoir les séquences SEQ ID N° 86,92,152,175,234,257,298,316,395,406,425,484 ;
- d) les polypeptides présentant au moins 80 % de préférence 85 %, 90 %, 95 % 30 et 98 % d'identité avec un polypeptide de a), b) ,c) ;
- e) les fragments biologiquement actifs des polypeptides de a), b), c), d)
- f) les polypeptides modifiés de a),b),c),d),e).

L'invention concerne également une séquence nucléotidique caractérisée en ce qu'elle comprend une séquence nucléotidique choisie parmi :

- a) une séquence nucléotidique telle que définie ci-dessus ;
- b) une séquence nucléotidique comportant au moins 80 % d'identité avec
5 une séquence nucléotidique de a) ;
- c) une séquence nucléotidique s'hybridant dans des conditions de forte stringence avec une séquence nucléotidique de a) ou b) ;
- d) une séquence nucléotidique complémentaire ou d'ARN correspondant à une séquence telle que définie en a), b) ou c) ;
- 10 e) une séquence nucléotidique de fragment représentatif d'une séquence telle que définie en a), b), c) ou d) ; et
- f) une séquence nucléotidique modifiée d'une séquence telle que définie en a), b), c), d) ou e).

Par acide nucléique, séquence nucléique ou d'acide nucléique, polynucléotide,
15 oligonucléotide, séquence de polynucléotide, séquence nucléotidique, termes qui seront employés indifféremment dans la présente description, on entend désigner un enchaînement précis de nucléotides, modifiés ou non, permettant de définir un fragment ou une région d'un acide nucléique, comportant ou non des nucléotides non naturels, et pouvant correspondre aussi bien à un ADN double brin, un ADN simple
20 brin que des produits de transcription desdits ADNs. Ainsi, les séquences nucléiques selon l'invention englobent également les PNA (Peptid Nucleic Acid), ou analogues.

Il doit être compris que la présente invention ne concerne pas les séquences nucléotidiques dans leur environnement chromosomique naturel, c'est-à-dire à l'état naturel. Il s'agit de séquences qui ont été isolées et/ou purifiées, c'est-à-dire qu'elles
25 ont été prélevées directement ou indirectement, par exemple par copie, leur environnement ayant été au moins partiellement modifié. On entend ainsi également désigner les acides nucléiques obtenus par synthèse chimique.

Par « pourcentage d'identité » entre deux séquences d'acides nucléiques ou d'acides aminés au sens de la présente invention, on entend désigner un pourcentage de
30 nucléotides ou de résidus d'acides aminés identiques entre les deux séquences à comparer, obtenu après le meilleur alignement, ce pourcentage étant purement statistique et les différences entre les deux séquences étant réparties au hasard et sur

toute leur longueur. On entend désigner par "meilleur alignement" ou "alignement optimal", l'alignement pour lequel le pourcentage d'identité déterminé comme ci-après est le plus élevé. Les comparaisons de séquences entre deux séquences d'acides nucléiques ou d'acides aminés sont traditionnellement réalisées en comparant ces séquences après les avoir alignées de manière optimale, ladite comparaison étant
5 réalisée par segment ou par « fenêtre de comparaison » pour identifier et comparer les régions locales de similarité de séquence. L'alignement optimal des séquences pour la comparaison peut être réalisé, outre manuellement, au moyen de l'algorithme d'homologie locale de Smith et Waterman (1981, Ad. App. Math. 2 : 482), au moyen
10 de l'algorithme d'homologie locale de Neddleman et Wunsch (1970, J. Mol. Biol. 48 : 443), au moyen de la méthode de recherche de similarité de Pearson et Lipman (1988, Proc. Natl. Acad. Sci. USA 85 : 2444), au moyen de logiciels informatiques utilisant ces algorithmes (GAP, BESTFIT, BLAST P, BLAST N, FASTA et TFASTA dans le Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr.,
15 Madison, WI). Afin d'obtenir l'alignement optimal, on utilise de préférence le programme BLAST, avec la matrice BLOSUM 62. On peut également utiliser les matrices PAM ou PAM250.

Le pourcentage d'identité entre deux séquences d'acides nucléiques ou d'acides aminés est déterminé en comparant ces deux séquences alignées de manière optimale
20 dans laquelle la séquence d'acides nucléiques ou d'acides aminés à comparer peut comprendre des additions ou des délétions par rapport à la séquence de référence pour un alignement optimal entre ces deux séquences. Le pourcentage d'identité est calculé en déterminant le nombre de positions identiques pour lesquelles le nucléotide ou le résidu d'acide aminé est identique entre les deux séquences, en divisant ce nombre de
25 positions identiques par le nombre total de positions comparées et en multipliant le résultat obtenu par 100 pour obtenir le pourcentage d'identité entre ces deux séquences.

Par séquences nucléiques présentant un pourcentage d'identité d'au moins 80 %, de préférence 85 % ou 90 %, de façon plus préférée 95 % voire 98 %, après alignement optimal avec une séquence de référence, on entend désigner les séquences nucléiques
30 présentant, par rapport à la séquence nucléique de référence, certaines modifications comme en particulier une délétion, une troncation, un allongement, une fusion chimérique et/ou une substitution, notamment ponctuelle, et dont la séquence nucléique

présente au moins 80 %, de préférence 85 %, 90 %, 95 % ou 98 %, d'identité après alignement optimal avec la séquence nucléique de référence. Il s'agit de préférence de séquences dont les séquences complémentaires sont susceptibles de s'hybrider spécifiquement avec les séquences de référence. De préférence, les conditions d'hybridation spécifiques ou de forte stringence seront telles qu'elles assurent au moins 80 %, de préférence 85 %, 90 %, 95 % ou 98 % d'identité après alignement optimal entre l'une des deux séquences et la séquence complémentaire de l'autre.

Une hybridation dans des conditions de forte stringence signifie que les conditions de température et de force ionique sont choisies de telle manière qu'elles permettent le maintien de l'hybridation entre deux fragments d'ADN complémentaires. A titre illustratif, des conditions de forte stringence de l'étape d'hybridation aux fins de définir les fragments polynucléotidiques décrits ci-dessus, sont avantageusement les suivantes.

L'hybridation ADN-ADN ou ADN-ARN est réalisée en deux étapes : (1) préhybridation à 42°C pendant 3 heures en tampon phosphate (20 mM, pH 7,5) contenant 5 x SSC (1 x SSC correspond à une solution 0,15 M NaCl + 0,015 M citrate de sodium), 50 % de formamide, 7 % de sodium dodécyl sulfate (SDS), 10 x Denhardt's, 5 % de dextran sulfate et 1 % d'ADN de sperme de saumon ; (2) hybridation proprement dite pendant 20 heures à une température dépendant de la taille de la sonde (i.e. : 42°C, pour une sonde de taille > 100 nucléotides) suivie de 2 lavages de 20 minutes à 20°C en 2 x SSC + 2 % SDS, 1 lavage de 20 minutes à 20°C en 0,1 x SSC + 0,1 % SDS. Le dernier lavage est pratiqué en 0,1 x SSC + 0,1 % SDS pendant 30 minutes à 60°C pour une sonde de taille > 100 nucléotides. Les conditions d'hybridation de forte stringence décrites ci-dessus pour un polynucléotide de taille définie, peuvent être adaptées par l'homme du métier pour des oligonucléotides de taille plus grande ou plus petite, selon l'enseignement de Sambrook et al., (1989, Molecular cloning : a laboratory manual. 2nd Ed. Cold Spring Harbor).

De plus, par fragment représentatif de séquences selon l'invention, on entend désigner tout fragment nucléotidique présentant au moins 15 nucléotides, de préférence au moins 20, 30, 75, 150, 300 et 450 nucléotides consécutifs de la séquence dont il est issu. On entend en particulier une séquence nucléique codant pour un fragment biologiquement actif d'un polypeptide, tel que défini plus loin,

notamment d'un polypeptide de séquence SEQ ID N°2 à 527.

Par fragment représentatif, on entend également les séquences intergéniques, et en particulier les séquences nucléotidiques portant les signaux de régulation (promoteurs, terminateurs, voire enhanceurs...).

5 Parmi lesdits fragments représentatifs, on préfère ceux ayant des séquences nucléotidiques correspondant à des cadres ouverts de lecture, dénommés séquences ORFs (ORF pour « Open Reading Frame »), compris en général entre un codon d'initiation et un codon stop, ou entre deux codons stop, et codant pour des polypeptides, de préférence d'au moins 30 acides aminés, tel que par exemple, sans
10 s'y limiter, les séquences ORFs qui seront décrites par la suite.

La numérotation des séquences nucléotidiques ORFs qui sera utilisée par la suite dans la présente description correspond à la numérotation des séquences d'acides aminés des protéines codées par lesdites ORFs.

Ainsi, les séquences nucléotidiques ORF2, ORF3..., ORF526 et ORF527
15 codent respectivement pour les protéines de séquences d'acides aminés SEQ ID N° 2, SEQ ID N° 3..., SEQ ID N° 526 et SEQ ID N° 527 figurant dans la liste de séquences de la présente invention. Les séquences nucléotidiques détaillées des séquences ORF2, ORF3..., ORF526 et ORF527 sont déterminées par leur position respective sur la séquence génomique SEQ ID N° 1 du cyanophage S2L. Le tableau 1
20 fournit les coordonnées de 54 ORFs préférées par rapport à la séquence nucléotidique SEQ ID N° 1, en donnant le nucléotide de départ, le nucléotide de fin d'ORF, et le cadre de lecture +1,2,3 ou -1,2,3 tel qu'expliqué ci-après. Le listing des séquences indique pour chacune des 526 ORF identifiées, numérotées ORF2 à ORF527, le cadre de lecture. Pour une parfaite concordance des numérotations des ORF et des
25 SEQ ID dans le listing de séquence, on a choisi de faire débiter les ORF au numéro 2 (il n'y a donc pas d'ORF 1). Il est entendu que la séquence SEQ ID N°1 est un brin ADN dans l'orientation 5'-3', la séquence SEQ ID N°2 est une séquence protéique codée par l'ORF N°2. Un cadre « positif » de +1 correspond au cadre de lecture désigné +1 commençant au nucléotide nt 3 de la SEQ ID N°1 (1^{er} codon de l'ORF2
30 situé sur ce cadre de lecture et commençant au nt 9 de SEQ ID N°1 : TCG qui correspond à la sérine S ; 2^{ème} codon de l'ORF 2 selon ce cadre : GAG qui correspond à l'acide glutamique E). Un cadre +2 correspond au cadre de lecture

désigné +2 et commençant au nucléotide nt 1 de la SEQ ID N°1 (1^{er} codon de l'ORF 4 situé sur ce cadre de lecture et commençant au nt 10 de SEQ ID N°1 : CGG qui correspond à l'arginine R ; 2^{ème} codon de l'ORF 4 selon ce cadre : AGG qui correspond à l'arginine R). Un cadre +3 correspond au cadre de lecture désigné +3
5 commençant au nucléotide nt 2 de la SEQ ID N°1 (1^{er} codon de l'ORF 5 situé sur ce cadre de lecture et commençant au nt35 de SEQ ID N°1 : CGT qui correspond à l'arginine R ; 2^{ème} codon de l'ORF 5 selon ce cadre : TCA qui correspond à la sérine S).

Ainsi l'ORF 2 commence au nt n°9 de SEQ ID n°1 (c'est à dire la base T) et
10 s'arrête au nt n°515 (c'est à dire la base G). L'ORF 4 commence au nt n°10 de SEQ ID n°1 (c'est à dire la base T) et s'arrête au nt n°342 (c'est à dire la base G). L'ORF 5 commence au nt n°35 de SEQ ID n°1 (c'est à dire la base C) et s'arrête au nt n°280 (c'est à dire la base A).

A l'inverse, un cadre négatif correspond au brin complémentaire antiparallèle
15 du brin positif. Par exemple pour une séquence ATG sur le brin positif dans le sens 5'-3', la séquence sur le brin complémentaire TAC se lira CAT. Par exemple pour l'ORF 3 (nt 9 à nt 791), le brin complémentaire des nucléotides 782 à 791 (CCT CGA TAG) est (GGA GCT ATC) se lisant dans le sens négatif CTA TCG AGG qui correspondent respectivement aux acides aminés L, S, R.

20 Les fragments représentatifs selon l'invention peuvent être obtenus par exemple par amplification spécifique telle que la PCR ou après digestion par des enzymes de restriction appropriés de séquences nucléotidiques selon l'invention, cette méthode étant décrite en particulier dans l'ouvrage de Sambrook et al.. Lesdits fragments représentatifs peuvent également être obtenus par synthèse chimique
25 lorsque leur taille n'est pas trop importante, selon des méthodes bien connues de l'homme du métier.

Parmi les séquences contenant des séquences de l'invention, ou des fragments représentatifs, on entend également les séquences qui sont naturellement encadrées par des séquences qui présentent au moins 80 %, 85 %, 90 %, 95 % ou 98 %
30 d'identité avec les séquences selon l'invention.

Par séquence nucléotidique modifiée, on entend toute séquence nucléotidique obtenue par mutagenèse selon des techniques bien connues de l'homme du métier, et

comportant des modifications par rapport aux séquences normales, par exemple des mutations dans les séquences régulatrices et/ou promotrices de l'expression du polypeptide, notamment conduisant à une modification du taux d'expression ou de l'activité dudit polypeptide.

- 5 Par séquence nucléotidique modifiée, on entend également toute séquence nucléotidique codant pour un polypeptide modifié tel que défini ci-après.

La présente invention fournit toutes les séquences nucléotidiques et polypeptidiques du génome du cyanophage S-2L. Par ailleurs, il est un objet de la présente invention que de divulguer les fonctions de ces gènes et protéines.

- 10 Les gènes décrits dans l'invention ont été isolés sur des fragments d'ADN grâce à des amorces déduites de la séquence du cyanophage S-2L.

- De manière préférée, l'invention est relative à une séquence nucléotidique caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans le métabolisme des nucléotides, des purines, des pyrimidines ou nucléosides. Dans ce texte, le terme « fragment représentatif»
15 pour un peptide signifie un fragment biologiquement actif de ce peptide (ayant une activité d'au moins 10, 20, 50, 100% de l'activité obtenue avec ce peptide).

- En particulier l'invention est relative à une séquence nucléotidique caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L ou un de ses fragments
20 représentatifs impliqué dans le métabolisme des nucléotides bases D, notamment un peptide de séquence SEQ ID N° 175 ou un de ses fragments représentatifs.

- De manière préférée, l'invention est relative à une séquence nucléotidique caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans le processus de réplication, notamment un
25 peptide de séquence SEQ ID N°14,18,142,355,429,454 ou un de leurs fragments représentatifs.

- De manière préférée, l'invention est relative à une séquence nucléotidique caractérisée en ce qu'elle code pour un polypeptide d'enveloppe, notamment de capsid, de cyanophage S-2L ou un de ses fragments représentatifs, notamment un
30 peptide de séquence SEQ ID N°169,316,351,392,395,406,422,425 ou un de leurs fragments représentatifs.

De manière préférée, l'invention est relative à une séquence nucléotidique selon l'invention caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L ou un de ses fragments impliqué dans le le détournement de la machinerie cellulaire.

5 De manière préférée, l'invention est relative à une séquence nucléotidique selon l'invention caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans le processus de transcription, notamment un peptide de séquence SEQ ID N°92,143,187,234 ou un de leurs fragments représentatifs.

10 De manière préférée, l'invention est relative à une séquence nucléotidique selon l'invention caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans le processus de virulence virale, notamment un peptide de séquence SEQ ID N° 257 ou un fragment représentatif.

15 De manière préférée, l'invention est relative à une séquence nucléotidique selon l'invention caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans les fonctions relatives aux transposons notamment un peptide de séquence SEQ ID N°208 ou un de ses fragments représentatifs.

20 Les fragments représentatifs de séquences nucléotidiques selon l'invention peuvent également être des sondes ou amorces, qui peuvent être utilisées dans des procédés de détection, d'identification, de dosage ou d'amplification de séquences nucléiques.

25 Une sonde ou amorce se définit, au sens de l'invention, comme étant un fragment d'acides nucléiques simple brin ou un fragment double brin dénaturé comprenant par exemple de 12 bases à quelques kb, notamment de 15 à quelques centaines de bases, de préférence de 15 à 50 ou 100 bases, et possédant une spécificité d'hybridation dans des conditions déterminées pour former un complexe d'hybridation avec un acide nucléique cible.

30 Les sondes et amorces selon l'invention peuvent être marquées directement ou indirectement par un composé radioactif ou non radioactif par des méthodes bien connues de l'homme du métier, afin d'obtenir un signal détectable et/ou quantifiable.

Les séquences de polynucléotides selon l'invention non marquées peuvent être utilisées directement comme sonde ou amorce.

Les séquences sont généralement marquées pour obtenir des séquences utilisables pour de nombreuses applications. Le marquage des amorces ou des sondes selon l'invention est réalisé par des éléments radioactifs ou par des molécules non radioactives.

Parmi les isotopes radioactifs utilisés, on peut citer le ^{32}P , le ^{33}P , le ^{35}S , le ^3H ou le ^{125}I . Les entités non radioactives sont sélectionnées parmi les ligands tels la biotine, l'avidine, la streptavidine, la dioxygénine, les haptènes, les colorants, les agents luminescents tels que les agents radioluminescents, chémoluminescents, bioluminescents, fluorescents, phosphorescents.

Les polynucléotides selon l'invention peuvent ainsi être utilisés comme amorce et/ou sonde dans des procédés mettant en oeuvre notamment la technique de PCR (amplification en chaîne par polymérase) (Rolfs et al., 1991, Berlin : Springer-Verlag). Cette technique nécessite le choix de paires d'amorces oligonucléotidiques encadrant le fragment qui doit être amplifié. On peut, par exemple, se référer à la technique décrite dans le brevet américain U.S. N° 4,683,202. Les fragments amplifiés peuvent être identifiés, par exemple après une électrophorèse en gel d'agarose ou de polyacrylamide, ou après une technique chromatographique comme la filtration sur gel ou la chromatographie échangeuse d'ions, puis séquencés. La spécificité de l'amplification peut être contrôlée en utilisant comme amorce les séquences nucléotidiques de polynucléotides de l'invention comme matrice, des plasmides contenant ces séquences ou encore les produits d'amplification dérivés. Les fragments nucléotidiques amplifiés peuvent être utilisés comme réactifs dans des réactions d'hybridation afin de mettre en évidence la présence, dans un échantillon biologique, d'un acide nucléique cible de séquence complémentaire à celle desdits fragments nucléotidiques amplifiés.

L'invention vise également les acides nucléiques susceptibles d'être obtenus par amplification à l'aide d'amorces selon l'invention.

D'autres techniques d'amplification de l'acide nucléique cible peuvent être avantageusement employées comme alternative à la PCR (PCR-like) à l'aide de couple d'amorces de séquences nucléotidiques selon l'invention. Par PCR-like on entend

désigner toutes les méthodes mettant en œuvre des reproductions directes ou indirectes des séquences d'acides nucléiques, ou bien dans lesquelles les systèmes de marquage ont été amplifiés, ces techniques sont bien entendu connues, en général il s'agit de l'amplification de l'ADN par une polymérase ; lorsque l'échantillon d'origine est un
5 ARN il convient préalablement d'effectuer une transcription reverse. Il existe actuellement de très nombreux procédés permettant cette amplification, comme par exemple la technique SDA (Strand Displacement Amplification) ou technique d'amplification à déplacement de brin (Walker et al., 1992, Nucleic Acids Res. 20 : 1691), la technique TAS (Transcription-based Amplification System) décrite par Kwoh
10 et al. (1989, Proc. Natl. Acad. Sci. USA, 86, 1173), la technique 3SR (Self-Sustained Sequence Replication) décrite par Guatelli et al. (1990, Proc. Natl. Acad. Sci. USA 87: 1874), la technique NASBA (Nucleic Acid Sequence Based Amplification) décrite par Kievitis et al. (1991, J. Virol. Methods, 35, 273), la technique TMA (Transcription Mediated Amplification), la technique LCR (Ligase Chain Reaction) décrite par
15 Landegren et al. (1988, Science 241, 1077), la technique de RCR (Repair Chain Reaction) décrite par Segev (1992, Kessler C. Springer Verlag, Berlin, New-York, 197-205), la technique CPR (Cycling Probe Reaction) décrite par Duck et al. (1990, Biotechniques, 9, 142), la technique d'amplification à la Q-béta-réplase décrite par Miele et al. (1983, J. Mol. Biol., 171, 281). Certaines de ces techniques ont depuis été
20 perfectionnées.

Dans le cas où le polynucléotide cible à détecter est un ARNm, on utilise avantageusement, préalablement à la mise en oeuvre d'une réaction d'amplification à l'aide des amorces selon l'invention ou à la mise en oeuvre d'un procédé de détection à l'aide des sondes de l'invention, une enzyme de type transcriptase inverse afin
25 d'obtenir un ADNc à partir de l'ARNm contenu dans l'échantillon biologique. L'ADNc obtenu servira alors de cible pour les amorces ou les sondes mises en oeuvre dans le procédé d'amplification ou de détection selon l'invention.

La technique d'hybridation de sondes peut être réalisée de manières diverses (Matthews et al., 1988, Anal. Biochem., 169, 1-25). La méthode la plus générale
30 consiste à immobiliser l'acide nucléique extrait des cellules de différents tissus ou de cellules en culture sur un support (tels que la nitrocellulose, le nylon, le polystyrène) et à incuber, dans des conditions bien définies, l'acide nucléique cible immobilisé avec la

sonde. Après l'hybridation, l'excès de sonde est éliminé et les molécules hybrides formées sont détectées par une méthode appropriée (mesure de la radioactivité, de la fluorescence ou de l'activité enzymatique liée à la sonde).

Selon un autre mode de mise en œuvre des sondes nucléiques selon l'invention, ces dernières peuvent être utilisées comme sondes de capture. Dans ce cas, une sonde, dite « sonde de capture », est immobilisée sur un support et sert à capturer par hybridation spécifique l'acide nucléique cible obtenu à partir de l'échantillon biologique à tester et l'acide nucléique cible est ensuite détecté grâce à une seconde sonde, dite « sonde de détection », marquée par un élément facilement détectable.

Parmi les fragments d'acides nucléiques intéressants, il faut ainsi citer en particulier les oligonucléotides anti-sens, c'est-à-dire dont la structure assure, par hybridation avec la séquence cible, une inhibition de l'expression du produit correspondant. Il faut également citer les oligonucléotides sens qui, par interaction avec des protéines impliquées dans la régulation de l'expression du produit correspondant, induiront soit une inhibition, soit une activation de cette expression.

De façon préférée, les sondes ou amorces selon l'invention sont immobilisées sur un support, de manière covalente ou non covalente. En particulier, le support peut être une puce à ADN ou un filtre à haute densité, également objets de la présente invention.

On entend désigner par puce à ADN ou filtre haute densité, un support sur lequel sont fixées des séquences d'ADN, chacune d'entre elles pouvant être repérée par sa localisation géographique. Ces puces ou filtres diffèrent principalement par leur taille, le matériau du support, et éventuellement le nombre de séquences d'ADN qui y sont fixées.

On peut fixer les sondes ou amorces selon la présente invention sur des supports solides, en particulier les puces à ADN, par différents procédés de fabrication. En particulier, on peut effectuer une synthèse in situ par adressage photochimique ou par jet d'encre. D'autres techniques consistent à effectuer une synthèse ex situ et à fixer les sondes sur le support de la puce à ADN par adressage mécanique, électronique ou par jet d'encre. Ces différents procédés sont connus de l'homme du métier.

Une séquence nucléotidique (sonde ou amorce) selon l'invention permet donc

la détection et/ou l'amplification de séquences nucléiques spécifiques. En particulier, la détection de cesdites séquences est facilitée lorsque la sonde est fixée sur une puce à ADN, ou à un filtre haute densité.

L'utilisation de puces à ADN ou de filtres à haute densité permet en effet de
5 déterminer l'expression de gènes dans un organisme présentant une séquence génomique proche du cyanophage S-2L.

La séquence génomique du cyanophage S-2L, complétée par l'identification de tous les gènes de cet organisme, telle que présentée dans la présente invention, sert de base à la construction de ces puces à ADN ou filtre.

10 La préparation de ces filtres ou puces consiste à synthétiser des oligonucléotides, correspondant aux extrémités 5' et 3' des gènes. Ces oligonucléotides sont choisis en utilisant la séquence génomique et ses annotations divulguées par la présente invention. La température d'appariement des ces oligonucléotides aux places correspondantes sur l'ADN doit être approximativement la même pour chaque oligonucleotide. Ceci permet
15 de préparer des fragments d'ADN correspondants à chaque gène par l'utilisation de condition de PCR appropriées dans un environnement hautement automatisé. Les fragments amplifiés sont ensuite immobilisés sur des filtres ou des supports en verre, silicium ou polymères synthétiques et ces milieux sont utilisés pour l'hybridation.

La disponibilité de tels filtres et/ou puces et de la séquence génomique
20 correspondante annotée permet d'étudier l'expression de grands ensembles, voire de la totalité des gènes dans des virus proches du cyanophage S-2L, en préparant les ADN complémentaires, et en les hybridant à l'ADN ou aux oligonucléotides immobilisés sur les filtres ou les puces. Egalement, les filtres et/ou les puces permettent d'étudier la variabilité des souches en préparant l'ADN de ces virus et en les hybridant à l'ADN ou
25 aux oligonucléotides immobilisés sur les filtres ou les puces.

Les différences entre les séquences génomiques des différentes souches ou espèces peuvent grandement affecter l'intensité de l'hybridation et, par conséquent, perturber l'interprétation des résultats. Il peut donc être nécessaire d'avoir la séquence précise des gènes de la souche que l'on souhaite étudier.

30 L'utilisation des filtres à haute densité et/ou des puces permet d'obtenir des connaissances nouvelles sur la régulation des gènes dans les organismes d'importance industrielle, et en particulier les bactéries recombinantes incorporant des gènes du

cyanophage S-2L propagées dans diverses conditions. Elle permet aussi une identification rapide des différences entre les génomes des souches utilisées dans de multiples applications industrielles.

Par ailleurs, les puces à ADN ou les filtres selon l'invention, contenant des
5 sondes ou amorces spécifiques du cyanophage S-2L, sont des éléments très avantageux de kits ou nécessaires pour la détection et/ou la quantification de l'expression de gènes du cyanophage S-2L dans des bactéries recombinantes intégrant ces gènes.

En effet, le contrôle de l'expression des gènes est un point critique pour les voies
10 métaboliques du cyanophage S-2L, soit en permettant l'expression d'un ou plusieurs gènes nouveaux, soit en modifiant l'expression de gènes déjà présents dans la cellule. La présente invention fournit l'ensemble des séquences naturellement actives chez le cyanophage S-2L permettant l'expression des gènes. Elle permet ainsi la détermination de l'ensemble des séquences exprimées chez le cyanophage S-2L. Elle fournit
15 également un outil permettant de repérer les gènes dont l'expression suit un schéma donné. Pour réaliser cela, l'ADN de tout ou partie des gènes du cyanophage S-2L peut être amplifié grâce à des amorces selon l'invention, puis fixé à un support comme par exemple le verre ou le nylon ou une puce à ADN, afin de construire un outil permettant de suivre le profil d'expression de ces gènes. Cet outil, constitué de ce support
20 contenant les séquences codantes sert de matrice d'hybridation à un mélange de molécules marquées reflétant les ARN messagers exprimés dans la cellule (en particulier les sondes marquées selon l'invention). En répétant cette expérience à différents instants et en combinant l'ensemble de ces données par un traitement approprié, on obtient alors les profils d'expression de l'ensemble de ces gènes. La
25 connaissance des séquences qui suivent un schéma de régulation donnée peut aussi être mise à profit pour rechercher de manière dirigée, par exemple par homologie, d'autres séquences suivant globalement, mais de manière légèrement différente le même schéma de régulation. En complément, il est possible d'isoler chaque séquence de contrôle présente en amont des segments servant de sondes et d'en suivre l'activité à
30 l'aide de moyen approprié comme un gène rapporteur (luciférase, β -galactosidase, GFP). Ces séquences isolées peuvent ensuite être modifiées et assemblées par ingénierie métabolique avec des séquences d'intérêt en vue de leur expression optimale.

La présente invention donne la liste de gènes codant ou susceptibles de coder pour des protéines régulant la transcription des gènes du cyanophage S-2L. Modifier la structure ou l'intégrité de ces gènes pourra permettre de modifier l'expression des gènes cibles contrôlés par des promoteurs cibles de ces régulateurs. Les indications données
 5 permettent de plus à l'homme du métier de choisir le ou les régulateurs pertinents pour l'application recherchée ainsi que leur cible, ce qui permet l'optimisation de l'expression de gènes d'intérêt. L'utilisation des outils précédemment décrits tels les puces à ADN, permet aussi de repérer l'ensemble des gènes dont la régulation est modifiée par cette inactivation. Il est ainsi possible de sélectionner un ensemble de
 10 séquence de contrôle répondant, à des nuances près, à un même type de régulation. Ces séquences peuvent être alors utilisées pour contrôler l'expression de gènes d'intérêt.

Selon un autre aspect, l'invention concerne les polypeptides comprenant :

- a) un polypeptide codé par une séquence nucléotidique selon l'invention telle que définie précédemment, en particulier un polypeptide codé par une ORF ;
- 15 b) un polypeptide présentant au moins 80 % de préférence 85 %, 90 %, 95 % et 98 % d'identité avec un polypeptide de a) ;
- c) un fragment biologiquement actif d'au moins 5, 7, 10 acides aminés d'un polypeptide selon a) ou b) ;
- d) un polypeptide modifié d'un polypeptide selon l'invention, ou tel que
 20 défini en a), b), ou c).

L'invention concerne de préférence :

- a) les polypeptides du cyanophage S-2L de séquences SEQ ID N°2 à SEQ ID N°527, codés respectivement par les ORF 2 à 527,
- b) les 54 polypeptides mentionnés sur le tableau 1 (SEQ ID
 25 N°14,18,26,68,86,92,105,109,134,142,143,148,152,169,175,187, 208,211,234,246,250,257,264,286,298,316,332,342,347,348, 351,355,364,365,369,370,392,395,406,418,422,425,429,432,433,454,464, 466,472,484,489,494,500.
- c) les 14 polypeptides du cyanophage S-2L, indiqués sur le tableau 1 comme
 30 présentant une homologie très significative, de séquence SEQ ID N° 86,92,152,175,234,257,298,316,395,406,425,484 .

L'invention concerne également :

- d) les polypeptides présentant au moins 80 % de préférence 85 %, 90 %, 95 % et 98 % d'identité avec un polypeptide de a), b), c)
- e) les fragments biologiquement actifs des polypeptides de a), b), c), d)
- f) les polypeptides modifiés de a), b), c), d), e).

5

L'invention concerne bien entendu tout spécialement les polypeptides impliqués dans la biosynthèse des bases D et des intermédiaires métaboliques de cette biosynthèse, en particulier le peptide de séquence SEQ ID N°175 à activité succidinylate synthétase.

- 10 Les phages pour lesquels les modifications sont pré-répliquatives, ce qui est vraisemblablement le cas du cyanophage S-2L, possèdent les séquences codantes de protéines nécessaires à la biosynthèse des bases modifiées, dans le cas présent des bases D. De plus, dans la mesure où les bases D font partie du génome du cyanophage, les enzymes polymérases notamment ADN polymérases doivent être
- 15 capables d'avoir la base D spécifiquement comme substrat au lieu de la base A. L'ADN polymérase du cyanophage S-2L est ainsi capable de discriminer le dDTP du dATP. De même, la transcription dépend d'une ARN polymérase spécifique et/ou d'un facteur sigma spécifique. L'invention concerne donc selon une réalisation préférée les polypeptides spécifiques à activité ADN polymérase, ARN polymérase
- 20 et facteurs associés, en particulier les peptides de séquence SEQ ID N° 92 et SEQ ID N°234 qui ont des activités spécifiques de la transcription d'ADN comprenant des bases D.

Dans la présente description, les termes polypeptides, séquences polypeptidiques, peptides et protéines sont interchangeable.

- 25 Il doit être compris que l'invention ne concerne pas les polypeptides sous forme naturelle, c'est-à-dire qu'ils ne sont pas pris dans leur environnement naturel mais qu'ils ont pu être isolés ou obtenus par purification à partir de sources naturelles, ou bien obtenus par recombinaison génétique, ou par synthèse chimique, et qu'ils peuvent alors comporter des acides aminés non naturels comme cela sera
- 30 décrit plus loin.

Par polypeptide présentant un certain pourcentage d'identité avec un autre, que l'on désignera également par polypeptide homologue, on entend désigner les

polypeptides présentant par rapport aux polypeptides naturels, certaines modifications, en particulier une délétion, addition ou substitution d'au moins un acide aminé, une troncation, un allongement, une solution chimérique et/ou une mutation, ou les polypeptides présentant des modifications post-traductionnelles.

- 5 Parmi les polypeptides homologues, on préfère ceux dont la séquence d'acides aminés présentent au moins 80 %, de préférence 85 %, 90 %, 95 % et 98 % d'homologie avec les séquences d'acides aminés des polypeptides selon l'invention. Dans le cas d'une substitution, un ou plusieurs acide(s) aminé(s) consécutif(s) ou non consécutif(s) sont remplacés par des acides aminés « équivalents ».
- 10 « acides aminés équivalents » vise ici à désigner tout acide aminé susceptible d'être substitué à l'un des acides aminés de la structure de base sans cependant modifier essentiellement les activités biologiques des peptides correspondant et telles qu'elles seront définies par la suite.

- Ces acides aminés équivalents peuvent être déterminés soit en s'appuyant sur
- 15 leur homologie de structure avec les acides aminés auxquels ils se substituent, soit sur des résultats d'essais comparatifs d'activité biologique entre les différents polypeptides susceptibles d'être effectués.

- A titre d'exemple, on mentionne les possibilités de substitution susceptibles d'être effectuées sans qu'il résulte en une modification approfondie de l'activité
- 20 biologique du polypeptide modifié correspondant. On peut remplacer ainsi la leucine par la valine ou l'isoleucine, l'acide aspartique par l'acide glutamine, la glutamine par l'asparagine, l'arginine par la lysine, etc... les substitutions inverses étant naturellement envisageables dans les mêmes conditions.

- Les polypeptides homologues correspondent également aux polypeptides codés
- 25 par les séquences nucléotidiques homologues ou identiques, telles que définies précédemment et comprennent ainsi dans la présente définition des polypeptides mutés ou correspondant à des variations inter ou intra espèces, pouvant exister chez le cyanophage S-2L, et qui correspondent notamment à des troncatures, substitutions, délétions et/ou additions, d'au moins un résidu d'acides aminés.

- 30 Il est entendu que l'on calcule le pourcentage d'identité entre deux polypeptides de la même façon qu'entre deux séquences d'acides nucléiques. Ainsi, le pourcentage d'identité entre deux polypeptides est calculé après alignement

optimal de ces deux séquences, sur une fenêtre d'homologie maximale. Pour définir ladite fenêtre d'homologie maximale, on peut utiliser les mêmes algorithmes que pour les séquences d'acide nucléique.

- Par fragment biologiquement actif d'un polypeptide selon l'invention, on entend désigner en particulier un fragment de polypeptide comportant au minimum 5 acides aminés, de préférence au moins 7, 10, 15, 25, 50, 75, 100, 150, 200, 250, 300 acides aminés, présentant au moins une des caractéristiques biologiques des polypeptides selon l'invention, notamment en ce qu'il est capable d'exercer de manière générale une activité même partielle, telle que par exemple :
- 10 - une activité enzymatique (métabolique) ou une activité pouvant être impliquée dans la biosynthèse ou la biodégradation de composés organiques ou inorganiques ;
 - une activité structurale (enveloppe cellulaire...) ;
 - une activité dans le processus de réplication, amplification, préparation, transcription, traduction ou maturation, notamment de l'ADN, de l'ARN
 - 15 ou des protéines
 - et tout particulièrement une activité impliquée dans la biosynthèse de bases D.

Les fragments de polypeptides peuvent correspondre à des fragments isolés ou purifiés naturellement présents dans les souches du cyanophage S-2L, ou à des fragments qui peuvent être obtenus par clivage dudit polypeptide par une enzyme protéolitique telle que la trypsine ou la chymotrypsine ou la collagénase, par un réactif chimique (bromure de cyanogène, CNBr) ou en plaçant ledit polypeptide dans un environnement très acide (par exemple à pH = 2,5). Des fragments polypeptidiques peuvent également être préparés par synthèse chimique, à partir d'hôtes transformés par un vecteur d'expression selon l'invention qui contiennent un acide nucléique permettant l'expression dudit fragment, et placé sous le contrôle des éléments de régulation et/ou d'expression appropriés.

Par « polypeptide modifié » d'un polypeptide selon l'invention, on entend désigner un polypeptide obtenu par recombinaison génétique ou par synthèse chimique comme décrit plus loin, qui présente au moins une modification par rapport à la séquence normale. Ces modifications peuvent être notamment portées sur des

acides aminés nécessaires pour la spécificité ou l'efficacité de l'activité, ou à l'origine de la conformation structurale, de la charge, ou de l'hydrophobicité du polypeptide selon l'invention. On peut ainsi créer des polypeptides d'activité équivalente, augmentée ou diminuée, ou de spécificité équivalente, plus étroite ou plus large. Parmi les polypeptides modifiés, il faut citer les polypeptides dans lesquels jusqu'à cinq acides aminés peuvent être modifiés, tronqués à l'extrémité N ou C-terminale, ou bien délétés, ou ajoutés.

Comme cela est indiqué, les modifications d'un polypeptide ont pour objectif notamment :

- 10 - de permettre sa mise en œuvre dans des procédés de biosynthèse ou de biodégradation de composés organiques ou inorganiques,
- de permettre sa mise en œuvre dans des procédés de réplication, d'amplification, de réparation et règle de transcription, de traduction, ou de maturation notamment de l'ADN, l'ARN, ou de protéines,
- 15 - de permettre sa sécrétion améliorée,
- de modifier sa solubilité, l'efficacité ou la spécificité de son activité, ou encore de faciliter sa purification.

La synthèse chimique présente également l'avantage de pouvoir utiliser des acides aminés non naturels ou des liaisons non peptidiques. Ainsi, il peut être intéressant d'utiliser des acides aminés non naturels, par exemple sous forme D, ou des analogues d'acides aminés, notamment des formes souffrées.

Sous un autre aspect, de manière préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il s'agit d'un polypeptide de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans le métabolisme des nucléotides, des purines, des pyrimidines ou nucléosides.

Sous un autre aspect, de manière préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il s'agit d'un polypeptide de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans le processus de réplication, et en ce qu'il est choisi parmi les polypeptides de séquence SEQ ID N°14,18,142,355,429,454 et un de leurs fragments.

L'invention concerne de manière très avantageuse des polypeptides de cyanophage S2L d'au moins 7 acides aminés et ayant une activité adénylosuccinate

synthetase. De manière préférée, de tels fragments comprennent le motif GSTGKG. En plus des résultats biologiques (métabolisme spécifique du cyanophage S2L capable de synthétiser et de polymériser de l'ADN incorporant des bases D), les inventeurs ont en effet identifié, des sites consensus en particulier les zones site de
5 liaison de phosphate et d'IMP. On retrouve en particulier le fragment QYGSTGKG, proche de la signature Prosite QWGDEGKG attribuée à l'adenylosuccinate synthétase, ou le fragment GSTGKG proche du fragment GDEGKG commun à *Escherichia coli*, *Methanobacterium thermoautotrophicum*, *Pyrococcus horikoshii* OT3. Les inventeurs ont identifié des homologues significatives pour les activités
10 adenylosuccinate synthétase, hélicase, facteur sigma, ces trois activités étant a priori directement liées étroitement avec le métabolisme spécifique des bases D.

Sous un autre aspect, de manière préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il s'agit d'un polypeptide de de cyanophage S-2L ou un de ses fragments impliqué dans le processus de transcription,
15 et en ce qu'il est choisi parmi les polypeptides de séquence SEQ ID N°92,143,187 et un de leurs fragments représentatifs.

Sous un autre aspect, de manière préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il s'agit d'un polypeptide d'enveloppe de cyanophage S-2L ou un de ses fragments, et en ce qu'il est choisi
20 parmi les polypeptides correspondant aux ORF169,316,351,392,395,406,422,425 et un de leurs fragments représentatifs.

Sous un autre aspect, de manière préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il s'agit d'un polypeptide de de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans le
25 détournement de la machinerie cellulaire ou dans le métabolisme intermédiaire.

Sous un autre aspect, de manière préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il s'agit d'un polypeptide de de cyanophage S-2L ou un de ses fragments représentatifs impliqué dans le processus de virulence, notamment le polypeptide de séquence SEQ ID N°247 et un de ses
30 fragments représentatifs.

Sous un autre aspect, de manière préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il s'agit d'un polypeptide de

cyanophage S-2L ou un de ses fragments impliqué dans les fonctions relatives aux transposons, notamment le polypeptide de séquence SEQ ID N°208 et un de ses fragments représentatifs.

5 Il faut noter toutefois qu'un organisme vivant est un tout et doit être pris comme tel. Ainsi, afin de pouvoir se développer et d'exhiber ses propriétés, tout organisme a besoin d'interactions entre les différentes voies métaboliques. Ainsi, la classification énoncée ci-dessus ne doit pas être considérée comme limitative, un gène pouvant être impliqué dans plusieurs voies métaboliques distinctes.

10 La présente invention a également pour objet les séquences nucléotidiques et/ou de polypeptides selon l'invention, caractérisées en ce que lesdites séquences sont enregistrées sur un support d'enregistrement dont la forme et la nature facilitent la lecture, l'analyse et/ou l'exploitation de ladite ou desdites séquence(s). Ces supports peuvent également contenir d'autres informations extraites de la présente
15 invention, notamment les analogies avec des séquences déjà connues, comme mentionné dans le Tableau 1 et/ou des informations concernant les séquences nucléotidiques et/ou de polypeptides d'autres microorganismes afin de faciliter l'analyse comparative et l'exploitation des résultats obtenus.

Parmi cesdits supports d'enregistrement, on préfère en particulier les supports
20 lisibles par un ordinateur, tels les supports magnétiques, optiques, électriques ou hybrides, en particulier les disquettes informatiques, les CD-ROM, les serveurs informatiques. De tels supports d'enregistrement sont également objet de l'invention.

Les supports d'enregistrement selon l'invention, avec les informations apportées, sont très utiles pour le choix d'amorces ou de sondes nucléotidiques pour
25 la détermination de gènes dans le cyanophage S-2L ou souches proches de cet organisme. De même, l'utilisation de ces supports pour l'étude du polymorphisme génétique de souche proche de du cyanophage S-2L, en particulier par la détermination des régions de colinéarité, est très utile dans la mesure où ces supports fournissent non seulement la séquence nucléotidique du génome du cyanophage S-
30 2L, mais également l'organisation génomique dans ladite séquence. Ainsi, les utilisations de supports d'enregistrement selon l'invention sont également des objets de l'invention.

Un procédé d'étude du polymorphisme génétique entre les souches proches du cyanophage S-2L, par détermination des régions de colinéarité, peut comprendre les étapes de

- 5 - fragmentation de l'ADN chromosomal de ladite autre souche (sonication, digestion),
- séquence des fragments d'ADN,
- analyse d'homologie avec le génome du cyanophage S-2L (SEQ ID N° 1).

10 Ce procédé qui comprend une étape d'analyse d'homologie avec le génome du cyanophage S-2L, en particulier grâce à l'aide d'un support d'enregistrement, est également l'objet de l'invention.

L'analyse d'homologie entre différentes séquences s'effectue en effet avantageusement à l'aide de logiciels de comparaisons de séquences, tels le logiciel Blast, ou les logiciels de la trousse GCG, décrits précédemment.

15 L'invention vise également une séquence nucléotidique telle que décrite précédemment, immobilisée sur un support, de manière covalente ou non-covalente, notamment un filtre à haute densité ou une puce à ADN.

L'invention vise également une séquence nucléotidique telle que décrite précédemment pour la détection et/ou l'amplification de séquences nucléiques.

20 Selon une réalisation un tel procédé de détection et d'amplification comprend par exemple les étapes suivantes :

- a) éventuellement, isolement de l'ADN à partir de l'échantillon biologique à analyser, ou obtention d'un ADNc à partir de l'ARN de l'échantillon biologique ;
- b) amplification spécifique de l'ADN du cyanophages S-2L à l'aide d'au
25 moins une amorce selon l'invention ;
- c) mise en évidence des produits d'amplification.

Ce procédé est basé sur l'amplification spécifique de l'ADN, en particulier par une réaction d'amplification en chaîne.

On préfère également un procédé comprenant les étapes suivantes :

- 30 a) mise en contact d'une sonde nucléotidique selon l'invention avec un échantillon biologique, l'acide nucléique contenu dans l'échantillon biologique ayant, le cas échéant, préalablement été rendu accessible à

l'hybridation, dans des conditions permettant l'hybridation de la sonde à l'acide nucléique du cyanophage S-2L ;

- b) mise en évidence de l'hybride éventuellement formé entre la sonde nucléotidique et l'ADN de l'échantillon biologique.

5 Un tel procédé ne doit pas être limité à la détection de la présence de l'ADN contenu dans l'échantillon biologique attesté, il peut être également mis en œuvre pour détecter l'ARN contenu dans ledit échantillon. Ce procédé englobe en particulier les Southern et Northern blot.

Ainsi, la présente invention englobe également un kit ou nécessaire pour la
10 détection et/ou l'identification de cyanophage S-2L, caractérisé en ce qu'il comprend les éléments suivants :

- a) une sonde nucléotidique selon l'invention ;
- b) éventuellement, les réactifs nécessaires à la mise en œuvre d'une réaction d'hybridation ;
- 15 c) éventuellement, au moins une amorce selon l'invention ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN.

De même, la présente invention englobe également les kits ou nécessaires pour la détection et/ou l'identification de cyanophage S-2L, comprenant les éléments suivants :

- 20 a) une sonde nucléotidique, dite sonde de capture, selon l'invention;
- b) une sonde oligonucléotidique, dite sonde de révélation, selon l'invention ;
- c) éventuellement, au moins une amorce selon l'invention ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN.

25 L'invention vise également les vecteurs de clonage et/ou d'expression, qui contiennent une séquence nucléotidique selon l'invention. On préfère en particulier les séquences nucléotidiques codant pour des polypeptides impliqués dans le métabolisme nucléotides, des purines, des pyrimidines ou nucléosides.

Les vecteurs selon l'invention comportent de préférence des éléments qui
30 permettent l'expression et/ou la sécrétion des séquences nucléotidiques dans une cellule hôte déterminée.

Le vecteur doit alors comporter un promoteur, des signaux d'initiation et de

terminaison de la traduction, ainsi que des régions appropriées de régulation de la transcription. Il doit pouvoir être maintenu de façon stable dans la cellule hôte et peut éventuellement posséder des signaux particuliers qui spécifient la sécrétion de la protéine traduite. Ces différents éléments sont choisis et optimisés par l'homme du
5 métier en fonction de l'hôte cellulaire utilisé. A cet effet, les séquences nucléotidiques selon l'invention peuvent être insérées dans des vecteurs à réplication autonome au sein de l'hôte choisi, ou être des vecteurs intégratifs de l'hôte choisi.

De tels vecteurs sont préparés par des méthodes couramment utilisées par l'homme du métier, et les clones résultant peuvent être introduits dans un hôte
10 approprié par des méthodes standard, telle que la lipofection, l'électroporation, le choc thermique, ou des méthodes chimiques.

Les vecteurs selon l'invention sont par exemple des vecteurs d'origine plasmidique ou virale. Ils sont utiles pour transformer des cellules hôtes afin de cloner ou d'exprimer les séquences nucléotidiques selon l'invention. On peut utiliser
15 comme vecteur directement le cyanophage S-2L lui-même.

L'invention comprend également les cellules hôtes transformées par un vecteur selon l'invention.

L'hôte cellulaire peut être choisi parmi des systèmes procaryotes ou eucaryotes, par exemple les cellules bactériennes mais également les cellules de
20 levure ou les cellules animales, en particulier les cellules de mammifères. On peut également utiliser des cellules d'insectes ou des cellules de plantes. Les cellules hôtes préférées selon l'invention sont en particulier les cellules procaryotes. Les cellules transformées selon l'invention sont utilisables dans des procédés de préparation de polypeptides recombinants selon l'invention.

Les procédés de préparation d'un polypeptide d'intérêt selon l'invention sous
25 forme recombinante, hors de l'environnement naturel, caractérisés en ce qu'ils mettent en œuvre un vecteur et/ou une cellule transformée par un vecteur selon l'invention sont eux-mêmes compris dans la présente invention. L'utilisation du cyanophage S-2L pour la production de tels peptides fait donc également partie de
30 l'invention.

De préférence, on cultive une cellule transformée par un vecteur selon l'invention dans des conditions qui permettent l'expression dudit polypeptide

d'intérêt et on récupère ledit peptide recombinant. Les cellules hôtes selon l'invention peuvent également être utilisées pour la préparation de compositions alimentaires, qui sont elles-mêmes objet de la présente invention.

Un tel procédé d'obtention de protéines d'intérêt du cyanophage S-2L, comprend selon une réalisation l'insertion de gènes d'intérêt du génome du phage S-2L, typiquement par ligation, dans des vecteurs de clonage et d'expression, dans des conditions permettant leur expression par la prise en charge par la machinerie de réplication d'un organisme hôte comme *E. coli*, et l'extraction des protéines produites.

Les messages héréditaires du phage recopiés sous forme d'ADN canonique sont aptes à s'exprimer comme des gènes de cyanobactéries. Les ARN messagers émis après réécriture de l'ADN de S-2L chez *E. coli* sont traduits en des protéines identiques à celles produites au cours de l'infection de *Synechococcus* par S-2L.

Selon une réalisation préférée les polypeptides d'intérêt sont des protéines impliquées dans le métabolisme de bases D, notamment la succinyladénylate synthétase.

En effet, il a été précisé plus haut qu'une base D est vraisemblablement formée par modification pré-répllicative et que des gènes cellulaires ont été recrutés à cette fin, deux voies de biosynthèse se présentant pour former le dDTP à partir d'un désoxynucléotide canonique, soit dAMP soit dGMP.

Suivant la première voie (figure 2a), le monomère activé dATP est d'emblée hydrolysé en dAMP par une enzyme du type de celle codée par *mutT* chez *E. coli* (9) ou du produit du gène *mutT* (9), ce qui a pour double effet de barrer l'accès de dATP à la synthèse de l'ADN et de fournir le précurseur de DMP. La biosynthèse de celui-ci s'effectue suivant les deux réactions successives convertissant IMP en GMP dans le métabolisme cellulaire (9) ; le nucléotide est finalement activé en dDTP en deux étapes de phosphorylation.

Suivant la deuxième voie (figure 2b), dDMP est obtenu en appliquant à dGMP les deux réactions convertissant dans les cellules IMP en AMP (9). Si elle prend aussi comme précurseur le dATP, cette deuxième voie est plus longue car dGMP doit préalablement être synthétisé via dIMP. Tout au long de cette deuxième voie, trois dNTP spécifiques et mutagènes sont formés (dIMP, dXMP et dSMP), contre un seul

(diGMP) dans la première (figure 2a).

Comme cela sera décrit plus loin les inventeurs ont réussi à identifier un ORF codant pour une enzyme de la seconde voie, la succinyladénylate synthétase.

Selon une autre réalisation préférée, les polypeptides d'intérêt sont des polymérase
5 de cyanophages S-2L, capables de polymériser des bases D, ce qui permet la propagation des acides nucléiques incorporant des bases D in vitro et in vivo.

Les inventeurs obtiennent en particulier des ADN polymérase spécialistes des duplex de haute stabilité et inaptes à répliquer dA pris comme constituant de la
10 matrice ou comme monomère triphosphate. Ces ADN polymérase seront typiquement obtenues par un procédé comprenant une étape d'expression, hors de l'environnement naturel, du gène de ladite ADN polymérase dans des bactéries recombinantes.

Selon une réalisation préférée les polypeptides d'intérêt sont des polypeptides
15 capables de modifier la transcription de l'ADN de cellules hôtes du cyanophage S-2L

En effet la transcription du génome de S-2L d'une ARN polymérase sur mesure, même si celle-ci n'est pas codée dans le phage. On sait que des enzymes de T4 altèrent l'ARN polymérase d'*E. coli*. Les promoteurs présents dans le génome de S-2L dévient du consensus connu de l'homme du métier (TATA box en particulier).
20 Il est vraisemblable que des facteurs d'amorçage de la transcription (sigma ou autre) soient codés ou modifiés par le phage, voire à ce qu'ils soient embarqués dans la capsid pour permettre le déclenchement du programme viral. Quoi qu'il en soit, le séquençage effectué par les inventeurs permet d'identifier sans effort excessif certains gènes de S-2L responsables du contrôle de la transcription par altération
25 chimique de l'ADN.

Ainsi qu'il a été dit, l'hôte cellulaire peut être choisi parmi des systèmes procaryotes ou eucaryotes. En particulier, il est possible d'identifier des séquences nucléotidiques selon l'invention, facilitant la sécrétion dans un tel système procaryote ou eucaryote. Un vecteur selon l'invention portant une telle séquence peut
30 donc être avantageusement utilisé pour la production de protéines recombinantes, destinées à être sécrétées. En effet, la purification de ces protéines recombinantes d'intérêt sera facilité par le fait qu'elles sont présentes dans le surnageant de la

culture cellulaire plutôt qu'à l'intérieur des cellules hôtes.

On peut également préparer les polypeptides selon l'invention par synthèse chimique. Un tel procédé de préparation est également un objet de l'invention. L'homme du métier connaît les procédés de synthèse chimique, par exemple les techniques mettant en œuvre des phases solides (voir notamment Steward et al., 1984, Solid phase peptides synthesis, Pierce Chem. Company, Rockford, 111, 2ème éd., (1984)) ou des techniques utilisant des phases solides partielles, par condensation de fragments ou par une synthèse en solution classique. Les polypeptides obtenus par synthèse chimique et pouvant comporter des acides aminés non naturels correspondant sont également compris dans l'invention.

L'invention comprend également les polypeptides hybrides qui comprennent au moins la séquence d'un polypeptide selon l'invention, et la séquence d'un polypeptide susceptible d'induire une réponse immunitaire chez l'homme ou l'animal. L'invention comprend également les séquences nucléotidiques qui codent pour de tels polypeptides hybrides, ou les vecteurs qui contiennent ces séquences nucléotidiques. Ce couplage entre un polypeptide selon l'invention et un polypeptide immunogène d'intérêt, peut être effectué par voie chimique, ou par voie biologique. Ainsi, selon l'invention, il est possible d'introduire un ou plusieurs élément(s) de liaison, notamment des acides aminés pour faciliter les réactions de couplage entre le polypeptide selon l'invention, et le polypeptide immunostimulateur, le couplage covalent de l'antigène immunostimulateur pouvant être réalisé à l'extrémité N ou C-terminale du polypeptide selon l'invention. Les réactifs bifonctionnels permettant ce couplage sont déterminés en fonction de l'extrémité choisie pour réaliser ce couplage, et les techniques de couplage sont bien connues de l'homme du métier.

Les conjugués issus d'un couplage de peptides peuvent être également préparés par recombinaison génétique. Le peptide hybride (conjugué) peut en effet être produit par des techniques d'ADN recombinant, par insertion ou addition à la séquence d'ADN codant pour le polypeptide selon l'invention, d'une séquence codant pour le ou les peptide(s) antigène(s), immunogène(s) ou haptène(s). Ces techniques de préparation de peptides hybrides par recombinaison génétique sont bien connues de l'homme du métier (voir par exemple Makrides, 1996, Microbiological Reviews 60,512-538).

De préférence, ledit polypeptide immunitaire est choisi dans le groupe des peptides contenant les anatoxines, notamment le toxoïde diphtérique ou le toxoïde tétanique, les protéines dérivées du Streptocoque (comme la protéine de liaison à la séralbumine humaine), les protéines membranaires OMPA et les complexes de
5 protéines de membranes externes, les vésicules de membranes externes ou les protéines de chocs thermiques.

Les séquences nucléotidiques et vecteurs, codant pour un polypeptide hybride selon l'invention sont également objet de l'invention.

Les polypeptides hybrides selon l'invention sont très utiles pour obtenir des
10 anticorps monoclonaux ou polyclonaux, capables de reconnaître spécifiquement les polypeptides selon l'invention. En effet, un polypeptide hybride selon l'invention permet la potentiation de la réponse immunitaire, contre le polypeptide selon l'invention couplé à la molécule immunogène. De tels anticorps monoclonaux ou polyclonaux, leurs fragments, ou les anticorps chimériques, reconnaissant les
15 polypeptides selon l'invention, sont également objets de l'invention.

Les anticorps monoclonaux spécifiques peuvent être obtenus selon la méthode classique de culture d'hybridome décrite par Köhler et Milstein (1975, Nature 256, 495).

Les anticorps selon l'invention sont par exemple des anticorps chimériques, des
20 anticorps humanisés, des fragments Fab, ou $F(ab')^2$. Il peut également se présenter sous forme d'immunoconjugué ou d'anticorps marqué afin d'obtenir un signal détectable et/ou quantifiable.

Les anticorps selon la présente invention sont notamment utilisables afin de détecter une expression d'un gène de cyanophage S-2L. En effet, la présence du
25 produit d'expression d'un gène reconnu par un anticorps spécifique dudit produit expression peut être détectée par la présence d'un complexe antigène-anticorps formé après la mise en contact d'une bactérie recombinante exprimant un gène d'intérêt donné du cyanophage S-2L avec un anticorps selon l'invention. La souche bactérienne utilisée peut avoir été « préparée », c'est-à-dire centrifugée, lysée, placée
30 dans un réactif approprié pour la constitution du milieu propice à la réaction immunologique. En particulier, on préfère un procédé de détection de l'expression d'un gène, correspondant à un Western blot, pouvant être effectué après une

électrophorèse sur gel de polyacrylamide d'un lysat de la souche bactérienne, en présence ou en l'absence de conditions réductrices (SDS-PAGE). Après migration et séparation des protéines sur le gel de polyacrylamide, on transfère lesdites protéines sur une membrane appropriée (par exemple en nylon) et on détecte la présence de la

5 protéine ou du polypeptide d'intérêt, par mise en contact de ladite membrane avec un anticorps selon l'invention.

Les polypeptides et les anticorps selon l'invention peuvent avantageusement être immobilisés sur un support, notamment une puce à protéines. Une telle puce à protéines est un objet de l'invention, et peut également contenir au moins un

10 polypeptide d'un microorganisme autre que le cyanophage S-2L ou un anticorps dirigé contre un composé d'un microorganisme autre que cyanophage S-2L.

Les puces à protéines ou filtres à haute densité contenant des protéines selon l'invention peuvent être construits de la même manière que les puces à ADN selon l'invention. En pratique, on peut effectuer la synthèse des polypeptides fixés

15 directement sur la puce à protéines, ou effectuer une synthèse ex situ suivie d'une étape de fixation du polypeptide synthétisé sur ladite puce. Cette dernière méthode est préférable, lorsque l'on désire fixer des protéines de taille importante sur le support, qui sont avantageusement préparées par génie génétique. Toutefois, si l'on ne désire fixer que des peptides sur le support de ladite puce, il peut être plus

20 intéressant de procéder à la synthèse desdits peptides directement in situ.

De préférence, on fixe un anticorps selon l'invention sur le support de la puce à protéines, et on détecte la présence de l'antigène correspondant, spécifique de Cyanophage S-2L ou d'un microorganisme associé.

Une puce à protéines ci-dessus décrite peut être utilisée pour la détection de

25 produits de gènes, pour établir un profil d'expression desdits gènes, en complément d'une puce à ADN selon l'invention.

Les puces à protéines selon l'invention sont également extrêmement utiles pour les expériences de protéomique, qui étudie les interactions entre les différentes protéines d'un microorganisme donné. De façon simplifiée, on fixe des peptides

30 représentatifs des différentes protéines d'un organisme sur un support. Puis, on met ledit support en contact avec des protéines marquées, et après une étape optionnelle de rinçage, on détecte des interactions entre lesdites protéines marquées et les

peptides fixés sur la puce à protéines.

Ainsi, les puces à protéines comprenant une séquence polypeptidique selon l'invention ou un anticorps selon l'invention sont objet de l'invention, ainsi que les kits ou nécessaires les contenant.

- 5 De préférence, les amorces et/ou sondes et/ou polypeptides et/ou anticorps selon la présente invention utilisés dans les procédés selon la présente invention sont choisis parmi les amorces et/ou sondes et/ou polypeptides et/ou anticorps spécifiques du Cyanophage S-2L

- 10 La présente invention a également pour objet les souches de Cyanophage S-2L et/ou de microorganismes associés contenant une ou plusieurs mutation(s) dans une séquence nucléotidique selon l'invention, en particulier une séquence ORF, ou leurs éléments régulateurs (en particulier promoteurs).

- On préfère, selon la présente invention, les souches de Cyanophage S-2L présentant une ou plusieurs mutation(s) dans les séquences nucléotidiques codant pour des polypeptides impliqués dans le métabolisme des bases D, la réplication et la transcription.

Lesdites mutations peuvent mener à une inactivation du gène, ou en particulier lorsqu'elles sont situées dans les éléments régulateurs dudit gène, à une surexpression de celui-ci.

- 20 Ainsi, on recherche en particulier des souches de Cyanophage S-2L sur-exprimant un polypeptide selon l'invention, impliquées dans les fonctions relatives à la synthèse de bases D ou de polynucléotides incorporant au moins une base D.

- L'art antérieur fait état de la connaissance du métabolisme spécifique du de cyanophage S-2L, conduisant à la synthèse de bases D au lieu de bases A. Toutefois, jusqu'à présent, sans connaître la séquence exacte du cyanophage S-2L, l'homme du métier n'avait pas à sa disposition les séquences ORF codantes et donc ne pouvait effectivement notamment cloner une séquence ORF donnée, tester l'activité biologique correspondante et exprimer des polypeptides d'intérêt. Ce type de procédé est désormais possible grâce au séquençage du génome du cyanophage S-2L effectué par les inventeurs.

Même sans connaître exactement à ce stade avec certitude la voie de synthèse des bases D, les inventeurs ont réussi à identifier des séquences codantes impliquées

dans cette voie métabolique. En testant successivement, mais sans effort excessif pour l'homme du métier, la fonction biologique des ORF susceptibles d'intervenir dans cette voie métabolique à partir des résultats obtenus (plus spécialement les ORF du groupe des polypeptides intervenant dans le métabolisme nucléotides, des purines, des pyrimidines ou nucléosides), les inventeurs peuvent donc repérer celles qui

5 codent pour les protéines déterminant cette voie.

L'invention concerne également selon un autre aspect l'utilisation des séquences polypeptidiques telles que décrites précédemment pour la production de bases D et ou de séquences polynucléotidiques comprenant des bases D. Ces

10 séquences polynucléotidiques seront notamment des séquences ADN, ou ARN, en particulier des ARNm.

L'invention concerne selon un autre aspect un procédé d'obtention de bases D et/ou de polynucléotides d'intérêt comprenant au moins une base D, ledit procédé comportant la culture d'un microorganisme contenant au moins une séquence

15 nucléotidique de cyanophage S-2L codante pour au moins un polypeptide impliqué dans la synthèse de bases D, dans des conditions appropriées pour le développement du vecteur et la synthèse de bases D. Typiquement le microorganisme cultivé comprend un vecteur tel que décrit précédemment contenant la ou lesdites séquences nucléotidiques de cyanophage S-2L codantes.

20 Selon une réalisation un tel procédé comprend :

- l'addition à un milieu comprenant les substrats nécessaires à l'obtention de bases D, d'un extrait ou mélange d'extraits de bactéries recombinantes exprimant au moins un gène de cyanophage S-2L impliqué dans la synthèse de bases D
- 25 - le cas échéant l'extraction de bases D et/ou de dits polynucléotides d'intérêt.

Selon une réalisation un tel procédé comprend :

- la préparation d'au moins une séquence ADN codante pour un polypeptide capable de provoquer chez un microorganisme hôte la
- 30 synthèse d'au moins une base D
- le clonage de ladite séquence codante dans un vecteur capable d'être transféré dans et de se répliquer dans ledit microorganisme hôte, ce

vecteur comprenant les éléments nécessaires à l'expression de ladite séquence codante

- le transfert du vecteur comprenant ladite séquence codante dans un microorganisme capable de produire les enzymes de la synthèse de bases D dirigée par ladite séquence codante
- la culture du microorganisme dans des conditions appropriées pour le développement du vecteur et la synthèse des bases D
- le cas échéant l'extraction de bases D et/ou de dits polynucléotides d'intérêt.

Comme cela sera décrit plus loin, les inventeurs ont réussi à cloner de l'ADN contenant des bases D, en utilisant des enzymes de restriction dont les sites de restriction ne possèdent pas de base A, notamment SmaI (site CCCGGG), SacII (site CCGCGG), MspI (site CCGG), BspRI (site GGCC). Les inventeurs ont montré que des enzymes de restriction comprenant au moins une base A n'hydrolysent pas l'ADN de S-2L : BamHI (GCATCC), EcoRI (GAATTC), HindIII (AAGCTT), Sau3AI (GATC). Pour cela les inventeurs sont allés à l'encontre d'un préjugé technique, à savoir que le clonage d'un ADN comprenant des bases D pourrait conduire à des ambiguïtés de copie lors du clonage. En effet, comme indiqué sur la figure 4, le clonage de « D DNA » chez *E.Coli*, est susceptible de conduire à des séquences différentes de celles issues du clonage de « A DNA ».

L'invention concerne selon un autre aspect un procédé d'obtention de bases D et/ou de polynucléotides d'intérêt comprenant au moins une base D ledit procédé comportant :

- l'addition, à un milieu comprenant les substrats nécessaires à l'obtention de bases D, du produit d'expression d'au moins un gène de cyanophage S-2L impliqué dans la synthèse de bases D, de manière à produire des bases D et/ou des polynucléotides d'intérêt comprenant au moins une base D
- le cas échéant l'extraction des bases D et/ou de dits polynucléotides d'intérêt.

Dans les procédés mentionnés ci-dessus, par synthèse de bases D et/ou de polynucléotides comprenant au moins une base D, on entend que les conditions de la synthèse sont telles que l'on obtienne uniquement ou essentiellement des bases D, ou

uniquement ou essentiellement des polynucléotides comprenant au moins une base D, ou à la fois des bases D et des polynucléotides comprenant au moins une base D dans des quantités souhaitées. Les quantités produites de bases D et de polynucléotides comprenant au moins une base D dépendront notamment du contrôle de l'expression de protéines impliquées dans la synthèses des bases D et dans l'incorporation des bases D lors de l'élongation des chaînes polynucléotidiques.

Selon un autre aspect l'invention concerne un procédé d'obtention de polynucléotides d'intérêt comprenant au moins une base D, ledit procédé comprenant la culture d'un microorganisme contenant au moins une séquence nucléotidique de cyanophage S-2L codante pour au moins un polypeptide impliqué dans l'élongation desdits polynucléotides avec incorporation de bases D, ADN polymérase notamment, dans des conditions appropriées pour le développement du vecteur et l'élongation desdits polynucléotides.

Selon un autre aspect l'invention concerne l'utilisation du cyanophage S-2L pour la production de réactifs utiles pour des réactions PCR ou PCR Like faisant intervenir des bases D. En particulier selon un mode de réalisation préféré ces réactifs seront des monomères dDTP.

En effet, le monomère dDTP est un bon substrat des ADN polymérases du cyanophage S-2L, et des matrices comportant la base D sont efficacement répliquées (1). La production biotechnologique de dD, dDMP et dDTP s'applique ainsi aux techniques de PCR, en augmentant la stabilité thermique des duplex, ou en masquant et en démasquant de nombreux sites de restriction (10). On entend que cette production n'est pas une production dans l'environnement naturel, la production dans l'environnement naturel signifiant une production par le cyanophage S-2L lui-même.

L'invention concerne également un procédé de production de polynucléotides d'intérêt comprenant au moins une base D, ledit procédé comportant une étape d'amplification, en présence de polymérase de cyanophage D et d'amorces appropriées, de polynucléotides comprenant au moins une base D.

Grâce à ce procédé, selon une technique de type PCR ou PCR like, à partir d'un polynucléotide d'intérêt comprenant au moins une base D de séquence connue, on obtient un grand nombre d'exemplaires de ce nucléotide.

Selon une réalisation le gène impliqué dans la synthèse de polynucléotides d'intérêt comprenant au moins une base D est le gène de la succinyladénylate synthétase. En effet, la succinyladénylate synthétase (ddba) catalyse la réaction de dGMP au dSMP lui-même transformé en dDMP (figure 2).

5 Selon une réalisation les polynucléotides d'intérêt sont des nucléosides d'intérêt thérapeutique.

Selon une réalisation, les polynucléotides d'intérêt sont produits par hémisynthèse ou par fermentation.

10 L'invention concerne en outre un procédé de sélection de composés capables de stimuler ou d'inhiber la synthèse de bases D et/ou de polynucléotides d'intérêt incorporant au moins une base D, comprenant l'addition au milieu de synthèse du composé testé et la comparaison de la synthèse en présence et en absence dudit composé.

15 L'invention concerne selon un autre aspect l'utilisation des séquences nucléotidiques de cyanophage S-2L telles que décrites précédemment pour tester leur fonction dans le métabolisme des nucléotides, des purines, des pyrimidines ou nucléosides, la réplication et la transcription.

20 Selon un autre aspect l'invention concerne l'utilisation du cyanophage S-2L pour la détermination de gènes permettant de réparer les mésappariements G :T ou iG :T survenant par désamination.

25 En effet, la base D elle-même est connue comme un mutagène chez *E. coli*. Ce fait pourrait s'expliquer par le fait que la désamination de D à la position 2 mène à l'isoguanine (iG), dont il a été récemment montré que le désoxynucléoside est mutagène (M. Bouzon, P. Marlière, résultats non publiés). La désamination de D à la position 6 mène à la guanine. Que cette dernière réaction de désamination survienne après incorporation de D dans l'ADN, elle résultera en une mutation au cycle de réplication suivant. Grâce au séquençage réalisé, l'identification de gènes aptes à réparer les mésappariements G:T ou iG:T survenant par désamination est désormais possible.

30 Selon un autre aspect l'invention concerne l'utilisation du cyanophage S-2L pour l'identification de gènes et la production de protéines capables de régénérer des 5'-termini.

En effet, la réplication de l'ADN du cyanophage, dont la stabilité est élevée (7,8), pourrait par ailleurs exiger des protéines auxiliaires (hélicase, SSB) sur mesure. Le génome est constitué d'un duplex linéaire, ce qui suppose une machinerie de régénération des 5'-termini, comme l'endonuclease utilisée pour résoudre les concatémères chez T7 (4), ou la protéine d'adduction en 5' chez phi29 (6), dont
5 l'activité pourrait nécessiter la présence de D dans leurs substrats.

Selon un autre aspect l'invention concerne l'utilisation du cyanophage S-2L pour l'identification de gènes capables de moduler l'activité des ribosomes

En effet, le cyanophage S-2L est par ailleurs susceptible de former un
10 précurseur ribonucléotidique portant la base D, pour le réduire ensuite en un désoxyribonucléotide correspondant, comme il advient pour les quatre bases de l'ARN (9). Dans ce cas, la transcription et la traduction des gènes phagiques pourraient s'effectuer via l'utilisation de codons, voire de tARNs comme chez T4 et T5, comportant cette base. Si une telle option a été empruntée par le phage, il est
15 possible que certains de ses gènes modulent l'activité des ribosomes.

Selon un autre aspect l'invention concerne l'utilisation du cyanophage S-2L pour l'identification ou la production de composés inhibiteurs de la biosynthèse des nucléotides puriques.

En effet les génomes phagiques spécifient toute une gamme d'inhibiteurs ayant
20 pour cible des enzymes cellulaires telles que la thymidylate synthase, la dUTPase, etc. (11). Dans le cas de S-2L, les inventeurs peuvent désormais identifier des inhibiteurs capables d'affecter la biosynthèse des nucléotides puriques.

L'invention concerne ainsi également un procédé utilisant de tels inhibiteurs pour contrôler le métabolisme ou l'expression génétique de cellules susceptibles
25 d'être infectées par un cyanophage S2L, notamment des cyanobactéries.

Dans la mesure où le contrôle du métabolisme des acides nucléiques ou nucléosides, en particulier des pyrimidines de l'ADN, est très utile en chimiothérapie et gérothérapie (2), l'invention concerne aussi un procédé utilisant de tels inhibiteurs pour contrôler ce métabolisme.

30

D'autres aspects et avantages de l'invention apparaîtront à la lecture de la description qui suit illustrée par les figures dans lesquelles :

- la figure 1 représente quelques exemples de bases modifiées
- les figures 2a et 2b représentent deux voies de biosynthèses possibles pour la synthèse de bases D par le cyanophage S-2L, la voie de la figure 2b étant la plus vraisemblable
- 5 - la figure 3 illustre schématiquement le génome du cyanophage S-2L
- la figure 4 représente schématiquement la difficulté potentielle du clonage de gènes incorporant des bases D chez *E. Coli*.

10 Des cyanophages S2L sont cultivés en masse à partir de l'espèce *Synechococcus elongatus* (8). L'ADN extrait est fragmenté par sonication pour constituer une banque shotgun clonée dans un vecteur chez *E. coli*. Les clones sont séquencés intensivement sur un séquenceur jusqu'à couverture complète du génome.

A mesure que des ORF sont élucidés comme homologues à des gènes connus,
15 il est procédé à leur expression chez *E. coli* ou chez *Synechococcus*, selon les fonctions supposées, notamment dans le but de valider les hypothèses fonctionnelles ou d'explorer les potentialités synthétiques.

A cette fin, les intermédiaires supposés des voies de synthèse (figure 2) ont été synthétisés suivant les méthodes courantes de la chimie des nucléosides et
20 nucléotides. Ils sont systématiquement soumis à des extraits ou des mélanges d'extraits de souches recombinantes exprimant chacune un gène de S-2L, pour identifier les activités enzymatiques spécifiées par le phage.

Plus précisément, l'ADN du phage S-2L a été préparé à partir de lysat de culture de *Synechococcus elongatus* en adaptant les techniques utilisées pour préparer
25 l'ADN du phage λ . Cet ADN a été digéré par différentes enzymes de restriction dont *SmaI*, ce qui a permis de vérifier que le profil de restriction obtenu était identique à celui décrit. Puis, on a montré que l'ADN de S-2L peut être répliqué chez *E. coli* et séquencé selon les protocoles standard, ce qui a conduit à construire une banque totale. Cette banque a été construite par insertion de fragments d'ADN digéré par
30 l'enzyme *CviJI* (de taille comprise entre 3 et 5 kb) dans le plasmide pBAM digéré par l'enzyme *SmaI* et déphosphorylé. Après électroporation de la souche d'*E. coli* DH10B, 400 clones ont été isolés et parmi ceux-ci 330 séquencés (290 dans les deux

orientations, 40 dans l'orientation + ou -) au Genoscope France. Les lectures ont été assemblées en un seul contig de 44,16 kb dont la composition en bases est conforme à celle de l'ADN du phage, c'est-à-dire 69,3 % G:C et 30,7% A:T (à la place de D:T). L'ensemble des ORFs déduites de ce contig a été comparé aux différentes banques de données, ce qui a permis d'annoter tout particulièrement 54 d'entre elles représentées sur le tableau 1 et tout particulièrement 14 d'entre elles (en ne tenant compte que des homologies statistiquement significatives avec des protéines bactériennes ou phagiques connues) mentionnées « très significatives » sur le tableau 1.

PROTEINE	Nombre aa	Cadre	Position	Très significatives	SEQ ID n° et ORF n°
ADN A réplication chromosomale initiateur de protéine	134	-2	661 – 1062		14
ADN Polymérase	50	1	963 – 1112		18
Polykétide synthase	177	-1	1308 – 1838		26
Beta glucosidase	135	-1	4698 – 5102		68
ADN hélicase	51	2	6280 – 6432	X	86
Sigma facteur	203	-3	6806 – 7414	X	92
Ribonucléoprotéine	100	1	8064 – 8363		105
ADN protéine de liaison	656	-2	8320 – 10287		109
ARN protéine de liaison	92	1	10299-10574		134
Réplique	55	-1	11052-11216		142
Putative RNA Pol IIIADN dirigée sous-unité large	72	3	11237-11452		143
DNA topoisomérase I	186	-1	11613-12170		148
ADN protéine d'emballage	448	-2	11956-13299	X	152
Protéine de capsid	109	1	13629-13955		169
Adenylosuccinate synthétase	419	-1	14235-15491	X	175
Putative transcriptase inverse	113	1	15132-15470		187
Transposase	65	3	16799-16993		208
Exodéoxyribonuclease VIII	347	-2	17113-18153	X	211
ADN hélicase	424	3	18962-20233	X	234
DSARN Adénosine désaminase	172	3	20237-20752		246
RRNA Adénine N-6 méthyltransférase	63	-2	20671-20859		250
Protéine de virulence E	738	3	20921-23134	X	257
ARN Polymérase II sous-unité large	69	-1	21444-21650		264
Putative ARNt endonucléase de clivage	251	-1	23316-24068		286
Enzyme de restriction de type I (M protéine)	60	-1	24072-24251	X	298
Protéine d'enveloppe	385	3	25685-26839	X	316
Guanylyl cyclase	111	-1	27183-27515		332
Uracyl-ADN glycosylase	65	-2	28063-28257		342
N-6 aminoadénine-N méthyltransférase	74	1	28239-28550		347

Inosine 5'monophosphate déshydrogénase	67	-1	28401-28601		348
Protéine de membrane	377	-1	28617-29747		351
Polymérase	94	-3	28871-29152		355
Transkétolese	82	-1	29751-29996		364
Ribulose biphosphate carboxylase	61	2	29893-30075		365
Protéine d'absorption de la queue	860	1	30105-32684		369
ARN protéine de liaison	247	2	30118-30858		370
ADN Pol III sous-unité gamma	103	-2	31876-32184		380
λ protéine de la queue M	159	1	32889-33365		392
λ protéine de la queue L	509	3	33023-34549	X	395
λ protéine de la queue K	271	2	30118-30858	X	406
ARN Pol bêta (fragment)	84	3	35267-35518		418
λ protéine de la queue I	236	2	35458-36165		422
λ protéine de la queue J	1456	1	35598-39965	X	425
ADN topoisomérase I	79	2	36169-36405		429
Dédoxyadénine méthylase	74	-3	36428-36649		432
ARN Polymérase II sous-unité large	146	2	36646-37083		433
ADN Polymérase I	63	-3	39089-39277		454
ADN gyrase sous-unité B	101	3	39989-40291		464
Dioxygénase	301	-2	40156-41058		466
ARN guanylyltransférase	190	1	41097-41666		472
Lysozyme	381	3	41951-43093	X	484
ARN protéine de liaison	129	-2	42457-42843		489
Tranposase/exonucléase	110	3	43097-43426		494
Phosphodiésterase	58	-2	43417-43590		500

Il s'agit notamment des protéines intervenant dans la formation et l'assemblage de la queue du bactériophage λ : protéine de la queue MLKI et J, de la protéine GP17 qui joue un rôle dans le packaging de l'ADN dans le bactériophage T4, d'une exonucléase qui pourrait intervenir dans l'exclusion de la base A, d'une RNA hélicase, d'un facteur sigma et d'une succinyladénylate synthétase.

L'identification de gènes codant pour un facteur sigma et une hélicase conduisent à conclure que la transcription du génome de S-2L et la répllication de l'ADN du cyanophage requièrent vraisemblablement des protéines spécifiques codées par le phage et dont l'activité pourrait dépendre de la base D.

D'autre part, il semble très vraisemblable que la base D soit formée par modification hémirépllicative. Entre les deux voies de biosynthèse de formation de dDTP décrites ci-dessus, l'identification d'un homologue du gène de la succinyladénylate synthétase appelée ddbA (pour désoxyribodiaminopurine biosynthetic gene A) conduisent à conclure que c'est la deuxième voie qui est

vraisemblablement empruntée lors de l'infection phagique (figure 2).

Plusieurs tests ont été effectués dans le but de déterminer l'activité de la protéine correspondante. Les résultats suggèrent que l'expression de *ddbA* permet de restaurer la croissance d'une souche d'*E. coli* exprimant le gène *yaaG* de *Bacillus subtilis* en présence d'une concentration élevée de dG (10mM). D'autre part, la 2,6-diaminopurine devient toxique (10mM) pour *E. coli* lorsqu'elle est sous forme phosphorylée (ce qui a été testé dans la même souche d'*E. coli* exprimant le gène *yaaG* de *Bacillus subtilis* soit MG1655 pSU *yaaG*) ce qui permet d'avoir un crible pour identifier in vivo la voie complète de biosynthèse de la base D.

- 10 Toutefois l'identification complète n'est pas nécessaire pour obtenir dès à présent des bases D par les procédés décrits précédemment.

Une autre approche consiste à exprimer systématiquement les ORF spécifiant tous les gènes possibles de S-2L et à combiner les activités brutes résultant de cette expression pour faire apparaître in vitro les métabolites de la voie. Une voie métabolique inductible produisant le dDTP sera ensuite édifiée chez *E. coli* par assemblage des gènes appropriés. La voie ainsi édifiée sera appliquée à des précurseurs synthétiques pour générer des nucléotides déviants par la base et le sucre.

- 15 L'utilisation de la base D dans les processus de réplication et de transcription est systématiquement recherchée dans les extraits des bactéries exprimant les ORF phagiques.

Les résultats ci-dessus ont été obtenus à l'aide des travaux suivants. Le gène *ddbA* a été exprimé chez *E. coli* sous le contrôle d'un promoteur inductible et plusieurs tests ont été effectués dans le but de déterminer l'activité de la protéine correspondante. Les résultats obtenus montrent que l'expression de *ddbA* permet de restaurer la croissance d'*E. coli* en présence d'une concentration élevée de dGMP. D'autre part, la 2,6-diaminopurine devient toxique pour *E. coli* lorsqu'elle est sous forme phosphorylée ce qui devrait permettre d'avoir un crible pour identifier in vivo la voie complète de biosynthèse de la base D. Le gène *ddbA* a été amplifié en utilisant 100 pmol de chaque oligonucléotide *ngaattcaagcttcagcgacggtagcgggcatac* et *nnnnccatggtgaagaactgcaacctgatc*, 100ng d'ADN de S-2L comme ADN matrice, 200mM de chaque dNTPs, 10 ml de tampon Pfu polymérase 10 fois concentré, DMSO 10% et 5U Pfu polymérase. Les cycles d'amplification étaient : une étape de

10 min à 95°C, puis 25 cycles 95°C 30 sec, 56°C 30 sec, 72°C 2 min 20 sec puis une étape de 10 min à 72°C. Le produit d'amplification a ensuite été purifié à l'aide du Kit Jetsorb (Genomed GmbH) puis digéré par les enzymes de restriction NcoI et HindIII. Après purification, le produit d'amplification a été inséré dans le plasmide pBAD24
5 (Guzman et al., 1995 J Bacteriol 177:4121-4130) digéré par les mêmes enzymes de restriction. Le gène ddbA est dans cette construction exprimé à partir du promoteur de l'opéron araBAD qui est inductible par l'arabinose.

La banque de cyanophage S2L est maintenue dans la souche d'*E.Coli* β 2033 déposée le 24 janvier 2001 à la Collection Nationale de Cultures de Microorganismes,
10 Institut Pasteur, 25 rue du Dr Roux, 75724 PARIS Cedex 15, France, selon les provisions du traité de Budapest, et a été enregistrée sous le numéro d'ordre I-2619.

Grâce aux travaux réalisés par les inventeurs, le séquençage du génome de S-2L permet d'altérer, d'inhiber ou de diversifier la synthèse des acides nucléiques in vitro et in vivo.

REVENDICATIONS

1. Séquence nucléotidique de cyanophage S-2L caractérisée en ce qu'elle correspond à SEQ ID N° 1.
2. Séquence nucléotidique de cyanophage S-2L, caractérisée en ce qu'elle est choisie parmi :
 - a) une séquence nucléotidique comportant au moins 80 % d'identité avec SEQ ID N° 1 ;
 - b) une séquence nucléotidique hybridant dans des conditions de forte stringence avec SEQ ID N° 1 ;
 - c) une séquence nucléotidique complémentaire de SEQ ID N° 1 ou complémentaire d'une séquence nucléotidique telle que définie en a), ou b), ou une séquence nucléotidique de l'ARN correspondant ;
 - d) une séquence nucléotidique de fragment représentatif de SEQ ID N° 1, ou de fragment représentatif d'une séquence nucléotidique telle que définie en a), b) ou c);
 - e) une séquence nucléotidique comprenant une séquence telle que définie en a), b), c) ou d) ; et
 - f) une séquence nucléotidique modifiée d'une séquence nucléotidique telle que définie en a), b), c), d) ou e).
3. Séquence nucléotidique selon la revendication 2, caractérisée en ce qu'elle code pour un polypeptide choisi parmi :
 - a) les polypeptides du cyanophage S-2L de séquences SEQ ID N°2 à SEQ ID N°527 ;
 - b) de préférence les polypeptides de séquence SEQ ID N°14,18,26,68,86,92,105,109,134,142,143,148,152,169,175,187, 208,211, 234,246, 250,257,264,286,298,316,332,342,347,348,351,355,364,365, 369,370,392,395, 406, 418,422,425,429,432,433,454,464,466,472, 484,489,494,500 ;
 - c) de préférence encore les polypeptides de séquence SEQ ID N° 86,92,152,175,

234,257,298,316,395,406,425,484 ;

- d) les polypeptides présentant au moins 80 % de préférence 85 %, 90 %, 95 % et 98 % d'identité avec un polypeptide de a), b) ,c)
- e) les fragments biologiquement actifs des polypeptides de a), b), c), d)
- 5 f) les polypeptides modifiés de a),b),c),d),e).

4. Séquence nucléotidique caractérisée en ce qu'elle comprend une séquence nucléotidique choisie parmi :

- a) une séquence nucléotidique selon la revendication 3 ;
- 10 b) une séquence nucléotidique comportant au moins 80 % d'identité avec une séquence nucléotidique selon la revendication 3 ;
- c) une séquence nucléotidique s'hybridant dans des conditions de forte stringence avec une séquence nucléotidique selon la revendication 3 ;
- d) une séquence nucléotidique complémentaire ou d'ARN correspondant à
- 15 une séquence telle que définie en a), b) ou c) ;
- e) une séquence nucléotidique de fragment représentatif d'une séquence telle que définie en a), b), c) ou d) ; et
- f) une séquence nucléotidique modifiée d'une séquence telle que définie en a), b), c), d) ou e).

20

5. Polypeptide codé par une séquence nucléotidique selon l'une des revendications 2 à 4.

- 6. Polypeptide selon la revendication 5, caractérisé en ce qu'il est choisi parmi les
- 25 peptides de séquence SEQ ID N°2 à 527, de préférence parmi les séquences SEQ ID N°14,18,26,68,86,92,105,109,134,142,143,148,152, 169,175,187, 208,211,234,246,250,257,264,286,298,316,332,342,347,348, 351,355,364,365, 369,370,392,395,406,418,422,425,429,432,433,454,464, 466,472,484,489,494,500.

30

7. Polypeptide selon la revendication 5 ou 6, caractérisé en ce qu'il est choisi parmi les séquences SEQ ID N° 86,92,152,175,234,257,298,316,395,406,425,

484.

8. Polypeptide caractérisé en ce qu'il comprend un polypeptide choisi parmi :
- 5 a) un polypeptide selon l'une des revendications 5 à 7 ;
 - b) un polypeptide présentant au moins 80 % d'identité avec un polypeptide selon l'une des revendications 5 à 7 ;
 - c) un fragment d'au moins 5 acides aminés d'un polypeptide selon l'une des revendications 5 à 7, ou tel que défini en b) ;
 - 10 d) un fragment biologiquement actif d'un polypeptide selon l'une des revendications 5 à 7, ou tel que défini en b) ou c) ; et
 - e) un polypeptide modifié d'un polypeptide selon l'une des revendications 5 à 7 ou tel que défini en b), c) ou d).
9. Séquence nucléotidique selon l'une des revendications 2 à 4, caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L impliqué dans la biosynthèse des nucléotides, des purines, des pyrimidines ou nucléosides ou pour l'un de ses fragments représentatifs.
10. Séquence nucléotidique selon l'une des revendications 2 à 4, caractérisée en ce qu'elle code pour un polypeptide de cyanophage S-2L impliqué dans la biosynthèse de bases D ou pour l'un de ses fragments représentatifs.
11. Séquence nucléotidique selon la revendication 10, caractérisée en ce qu'elle
- 25 code pour un peptide de séquence SEQ ID N°175.
12. Séquence nucléotidique selon l'une des revendications 2 à 4, caractérisée en ce qu'elle code pour un polypeptide de Cyanophage S-2L impliqué dans le processus de réplication, ou pour l'un de ses fragments représentatifs, de
- 30 préférence un peptide de séquence SEQ ID N°14,18,142,355,429,454 (activité ADN polymérase, topoisomérase).

13. Séquence nucléotidique selon l'une des revendications 2 à 4, caractérisée en ce qu'elle code pour un polypeptide de Cyanophage S-2L impliqué dans le processus de transcription ou pour l'un de ses fragments représentatifs, de préférence un peptide de séquence SEQ ID N°92,143,187,234 encore de préférence SEQ ID N°92.
14. Séquence nucléotidique selon l'une des revendications 2 à 4, caractérisée en ce qu'elle code pour un polypeptide d'enveloppe de Cyanophage S-2L ou l'un de ses fragments représentatifs, de préférence un peptide de séquence SEQ ID N°169,316,351,392,395,406,422,425, notamment un peptide de séquence SEQ ID N°395,406,425.
15. Séquence nucléotidique selon l'une des revendications 2 à 4, caractérisée en ce qu'elle code pour un polypeptide de Cyanophage S-2L impliqué dans le détournement de la machinerie cellulaire ou l'un de ses fragments représentatifs.
16. Séquence nucléotidique selon l'une des revendications 2 à 4, caractérisée en ce qu'elle code pour un polypeptide de Cyanophage S-2L impliqué dans la virulence ou l'un de ses fragments représentatifs, de préférence un peptide de séquence SEQ ID N°257.
17. Polypeptide selon l'une des revendications 5 à 8, caractérisé en ce qu'il s'agit d'un polypeptide de Cyanophage S-2L impliqué dans la biosynthèse des nucléotides, des purines, des pyrimidines ou nucléosides.
18. Polypeptide selon l'une des revendications 5 à 8, caractérisé en ce qu'il s'agit d'un polypeptide de Cyanophage S-2L impliqué dans la biosynthèse de bases D, de préférence un peptide de séquence SEQ ID N°175 ou un fragment représentatif.
19. Polypeptide selon l'une des revendications 5 à 8, caractérisé en ce qu'il s'agit

d'un polypeptide de Cyanophage S-2L impliqué dans le processus de réplication, de préférence un peptide de séquence SEQ ID N°14,18,142,355,429,454 ou un fragment représentatif.

- 5 20. Polypeptide selon l'une des revendications 5 à 8, caractérisé en ce qu'il s'agit d'un polypeptide de Cyanophage S-2L impliqué dans le processus de transcription, de préférence un peptide de séquence SEQ ID N°92,143,187 ou un fragment représentatif.
- 10 21. Polypeptide selon l'une des revendications 5 à 8, caractérisé en ce qu'il s'agit d'un polypeptide d'enveloppe de Cyanophage S-2L, de préférence un peptide de séquence SEQ ID N°169,316,351,392,395,406,422,425 ou un fragment représentatif.
- 15 22. Polypeptide selon l'une des revendications 5 à 8, caractérisé en ce qu'il s'agit d'un polypeptide de Cyanophage S-2L impliqué dans le métabolisme intermédiaire, ou un fragment représentatif.
- 20 23. Polypeptide selon l'une des revendications 5 à 8, caractérisé en ce qu'il s'agit d'un polypeptide de Cyanophage S-2L impliqué dans le processus de détournement de la machinerie cellulaire ou l'un de ses fragments.
- 25 24. Puce à ADN ou filtre, caractérisée en ce qu'elle contient au moins une séquence nucléotidique selon l'une quelconque des revendications 2 à 4.
- 25 25. Vecteur de clonage, et/ou d'expression, caractérisé en ce qu'il contient une séquence nucléotidique selon l'une des revendications à 3 ou 4 ou 9 à 16.
- 30 26. Vecteur de clonage, et/ou d'expression selon la revendication 25, caractérisé en ce qu'il contient une séquence nucléotidique selon la revendication 10 ou 11, en particulier codant pour une protéine impliquée dans la synthèse de bases D.

27. Vecteur de clonage, et/ou d'expression selon la revendication 25, caractérisé en ce qu'il contient une séquence nucléotidique selon la revendication 11 ou 12, en particulier codant pour une polymérase capable de polymériser des bases D.
- 5 28. Cellule hôte, caractérisée en ce qu'elle est transformée par un vecteur selon l'une des revendications 25 à 27.
29. Cellule hôte selon la revendication 28, caractérisée en ce qu'il s'agit d'une bactérie.
- 10 30. Procédé de préparation d'un polypeptide d'intérêt, caractérisé en ce que l'on cultive une cellule transformée par un vecteur selon l'une des revendications 25 à 27 dans des conditions permettant l'expression dudit polypeptide et que l'on recupère ledit polypeptide recombinant.
- 15 31. Procédé selon la revendication 30, caractérisé en ce que le polypeptide d'intérêt est une protéine impliquée dans le métabolisme de bases D, notamment la succinyladénylate synthétase.
- 20 32. Procédé selon la revendication, caractérisé en ce que le polypeptide d'intérêt est une polymérase de cyanophages S-2L, capable de polymériser des bases D.
33. Polypeptide recombinant susceptible d'être obtenu par un procédé selon la revendication 30.
- 25 34. Anticorps monoclonal ou polyclonal, ses fragments, ou anticorps chimérique, caractérisé en ce qu'il est capable de reconnaître spécifiquement un polypeptide selon l'une des revendications 5 à 8 ou 17 à 23.
- 30 35. Anticorps selon la revendication 34, caractérisé en ce qu'il s'agit d'un anticorps marqué.

36. Puce à protéine, caractérisée en ce qu'elle contient au moins un polypeptide selon l'une des revendications 5 à 8 ou 17 à 23, ou au moins un anticorps selon l'une des revendications 34 ou 35, immobilisé sur le support de ladite puce.
- 5
37. Procédé de détection et/ou d'identification du cyanophage S-2L ou d'un phage associé dans un échantillon biologique, caractérisé en ce qu'il met en œuvre une séquence nucléotidique selon l'une des revendications 3 à 4 ou 9 à 16.
- 10
38. Procédé d'obtention de bases D et/ou de polynucléotides d'intérêt comprenant au moins une base D, comportant la culture d'un microorganisme comprenant au moins une séquence nucléotidique de cyanophage S-2L codante pour au moins un polypeptide impliqué dans la synthèse de bases D, dans des conditions appropriées pour le développement du vecteur et la synthèse de bases D et/ou dedits polynucléotides d'intérêt.
- 15
39. Procédé selon la revendication 38 comportant :
- l'addition à un milieu comprenant les substrats nécessaires à l'obtention de bases D, d'un extrait ou un mélange d'extrait de bactéries recombinantes exprimant au moins un gène de cyanophage S-2L impliqué dans la synthèse de bases D
 - le cas échéant l'extraction des polynucléotides d'intérêt.
- 20
40. Procédé selon la revendication 38 comportant :
- la préparation d'au moins une séquence ADN codante pour un polypeptide capable de provoquer chez un microorganisme hôte la synthèse d'au moins une base D
 - le clonage de ladite séquence codante dans un vecteur capable d'être transféré dans et de se répliquer dans ledit microorganisme hôte, ce vecteur comprenant les éléments nécessaires à l'expression de ladite séquence codante
- 25
- 30

- le transfert du vecteur comprenant ladite séquence codante dans un microorganisme capable de produire les enzymes de la synthèse de bases D dirigée par ladite séquence codante
 - la culture du microorganisme dans des conditions appropriées pour le développement du vecteur et la synthèse des bases D
 - le cas échéant l'extraction de bases D et/ou dedits polynucléotides d'intérêt.
41. Procédé d'obtention de bases D et/ou de polynucléotides d'intérêt comprenant au moins une base D ledit procédé comportant :
- l'addition, à un milieu comprenant les substrats nécessaires à l'obtention de bases D, du produit d'expression d'au moins un gène de cyanophage S-2L impliqué dans la synthèse de bases D, de manière à produire des bases D et/ou des polynucléotides d'intérêt comprenant au moins une base D
 - l'extraction des bases D et/ou dedits polynucléotides d'intérêt.
42. Procédé d'obtention de polynucléotides d'intérêt comprenant au moins une base D, ledit procédé comprenant la culture d'un microorganisme contenant au moins une séquence nucléotidique de cyanophage S-2L codante pour au moins un polypeptide impliqué dans l'élongation desdits polynucléotides avec incorporation de bases D, ADN polymérase notamment, dans des conditions appropriées pour le développement du vecteur et l'élongation desdits polynucléotides.
43. Procédé selon l'une quelconque des revendications 38 à 42 caractérisé en ce que le gène impliqué dans la synthèse de bases D est le gène de la succinyladénylate synthétase.
44. Procédé selon l'une quelconque des revendications 38 à 42 caractérisé en ce qu'il comprend une étape d'amplification, en présence de polymérase de cyanophage D et d'amorces appropriées, de polynucléotides comprenant au moins une base D.

45. Procédé de sélection de composés capables de stimuler ou d'inhiber la synthèse de bases D et/ou de polynucléotides d'intérêt comprenant au moins une base D, comprenant l'addition au milieu de synthèse du composé testé et la comparaison de la synthèse en présence et en absence dudit composé.

5

46. Utilisation du cyanophage S-2L pour l'obtention d'ADN polymérase ou d'ARN polymérase impliquées dans le métabolisme des bases D.

47. Utilisation du cyanophage S-2L pour la fabrication de nucléotides d'intérêt non
10 obtenus naturellement comprenant au moins une base D, dDMP et dDTP.

48. Souche de Cyanophage S-2L déposée à la CNCM n° I-2619 caractérisée en ce qu'elle contient au moins une séquence nucléotidique selon la revendication 1 à 4,

15 49. Procédé de production d'une banque génomique de cyanophage S2L caractérisée en ce que le procédé comprend les étapes suivantes :

- a) Culture des cyanophages S2L purifié à partir de l'espèce *Synechococcus*,
- b) Fragmentation de l'ADN extrait par la technique de sonication,
- c) Clonage de la banque Shotgun obtenue dans un vecteur *E.coli*, et
- 20 d) Séquençage des clones jusqu'à couverture complète du génome.

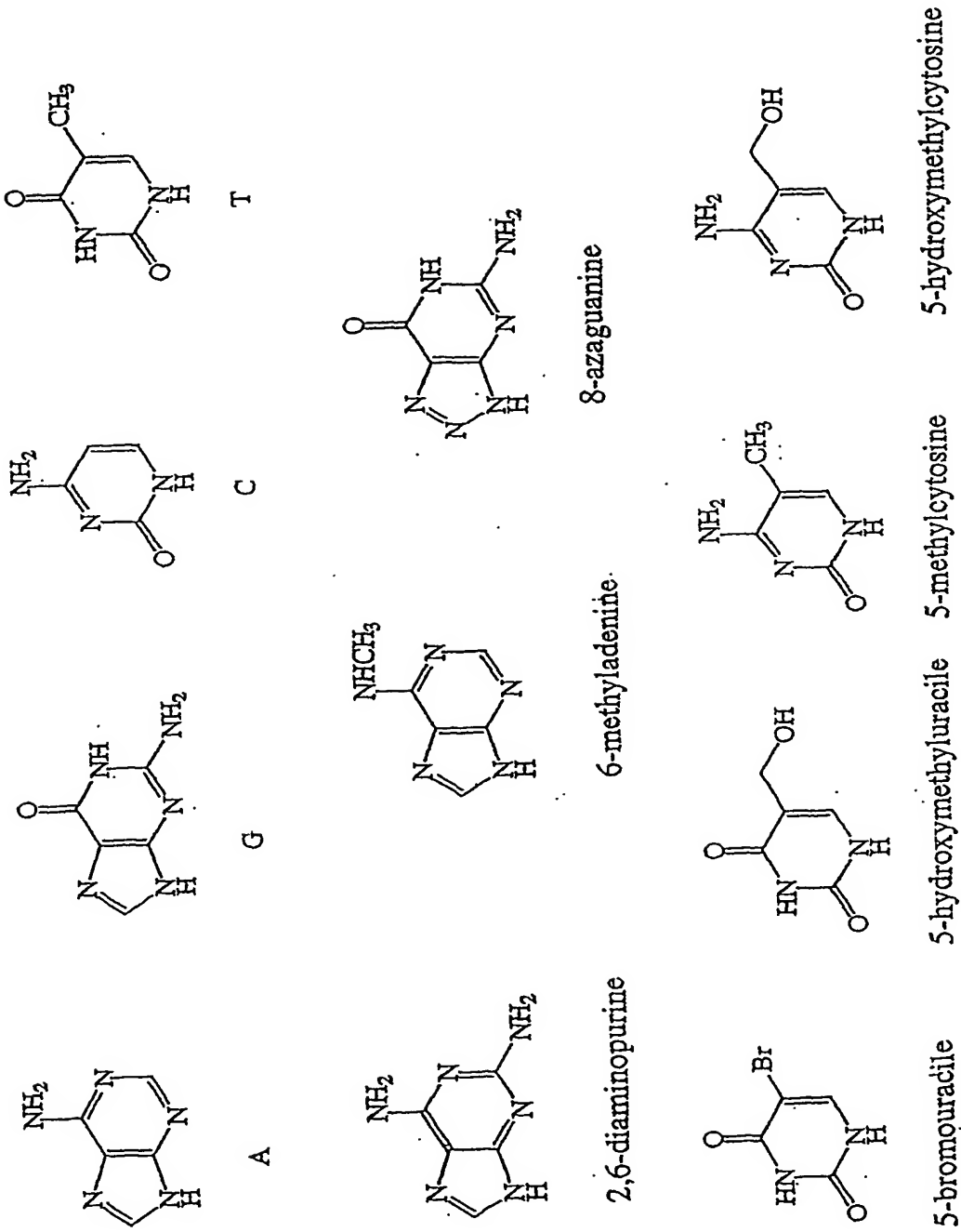
50. Banque génomique de cyanophage S2L déposée le 24 janvier 2001 à la C.N.C.M. selon les provisions du traité de Budapest, et enregistrées sous le numéro d'accession I-2619 obtenue par un procédé selon la revendication 49.

25

51. Plasmides contenus dans des bactéries recombinantes déposées à la C.N.C.M. sous la référence I-2619.

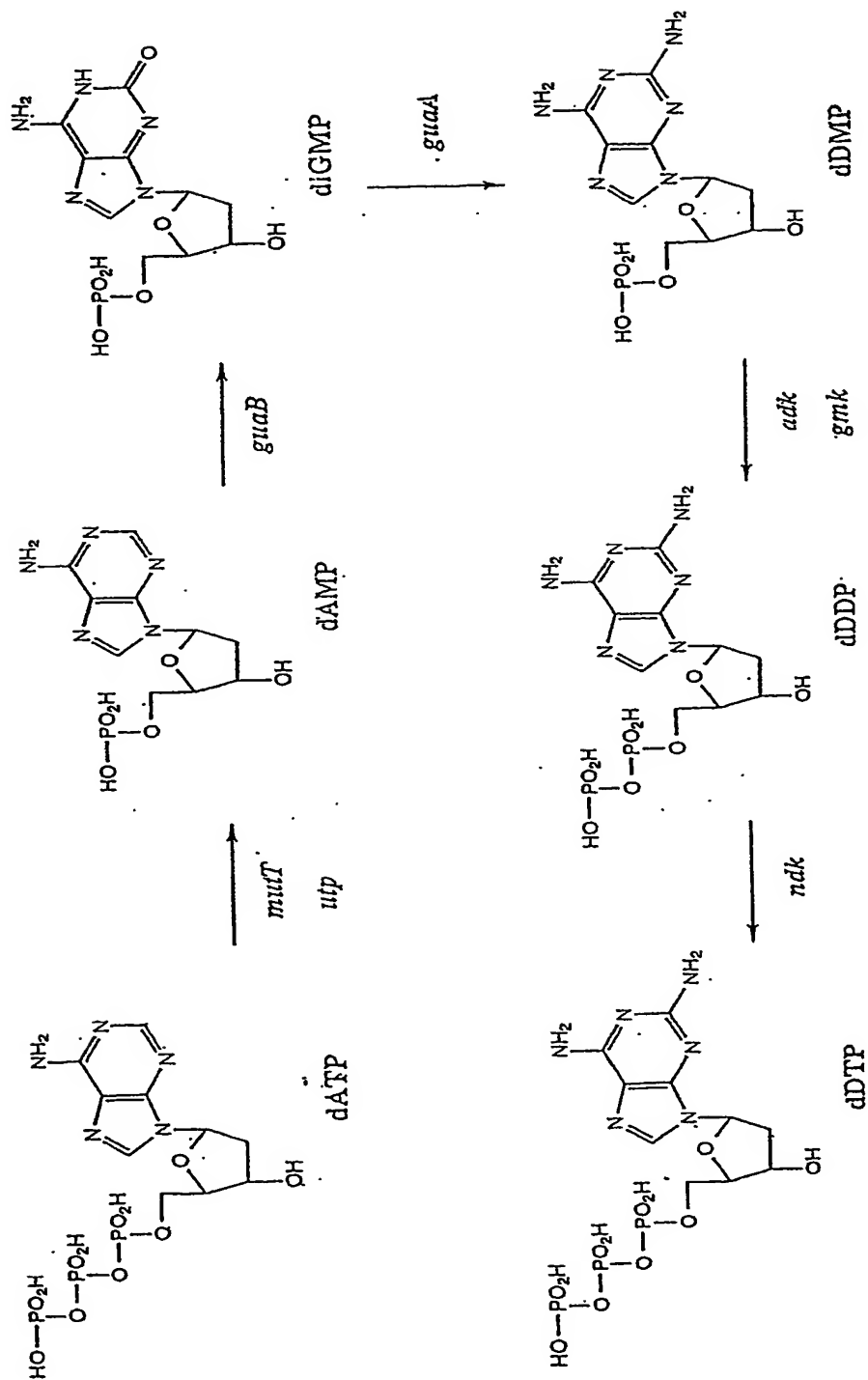
52. Bactéries recombinantes telles que déposées à la C.N.C.M. sous la référence
30 I-2619.

Figure 1



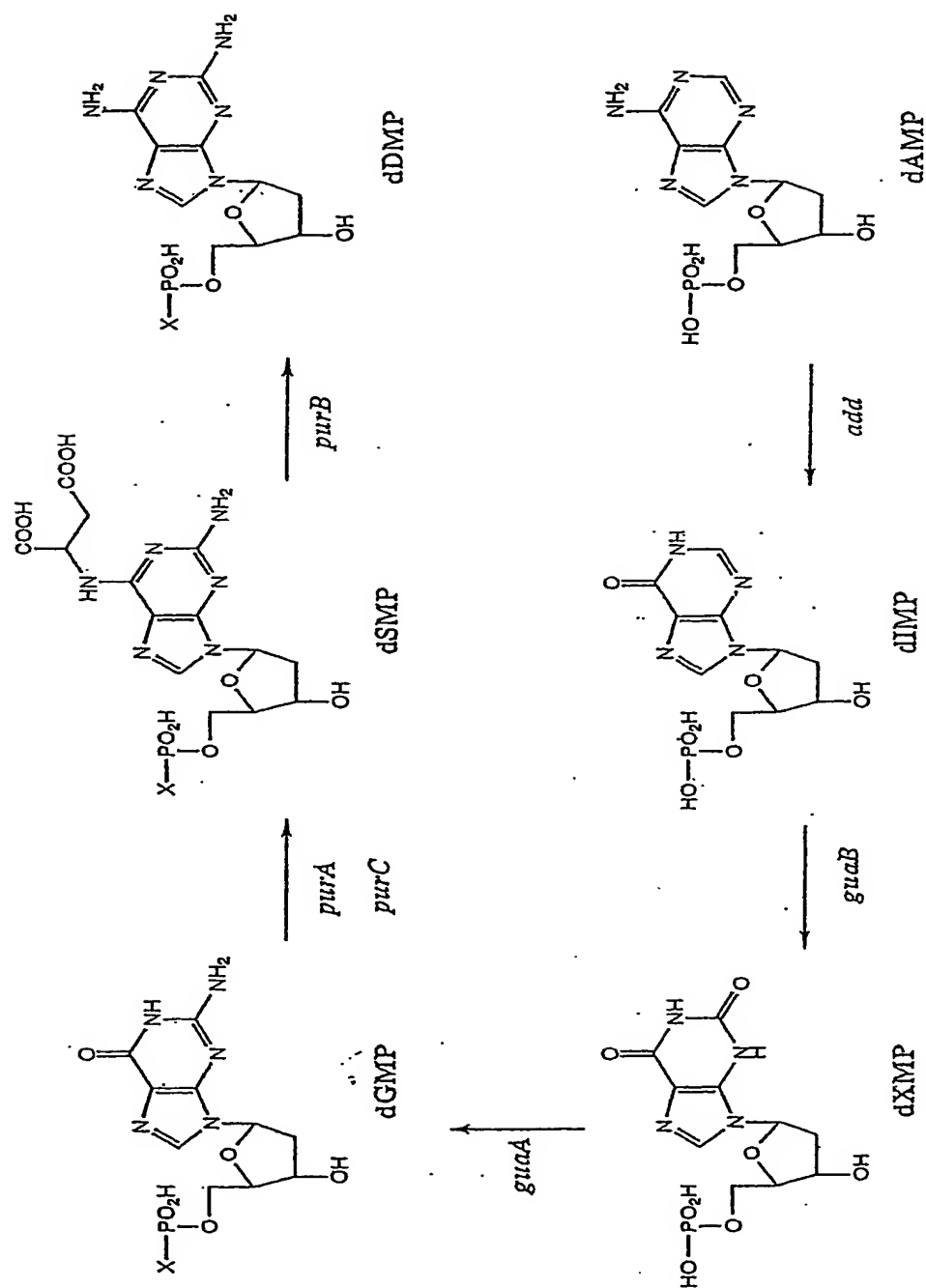
2/5

Figure 2a



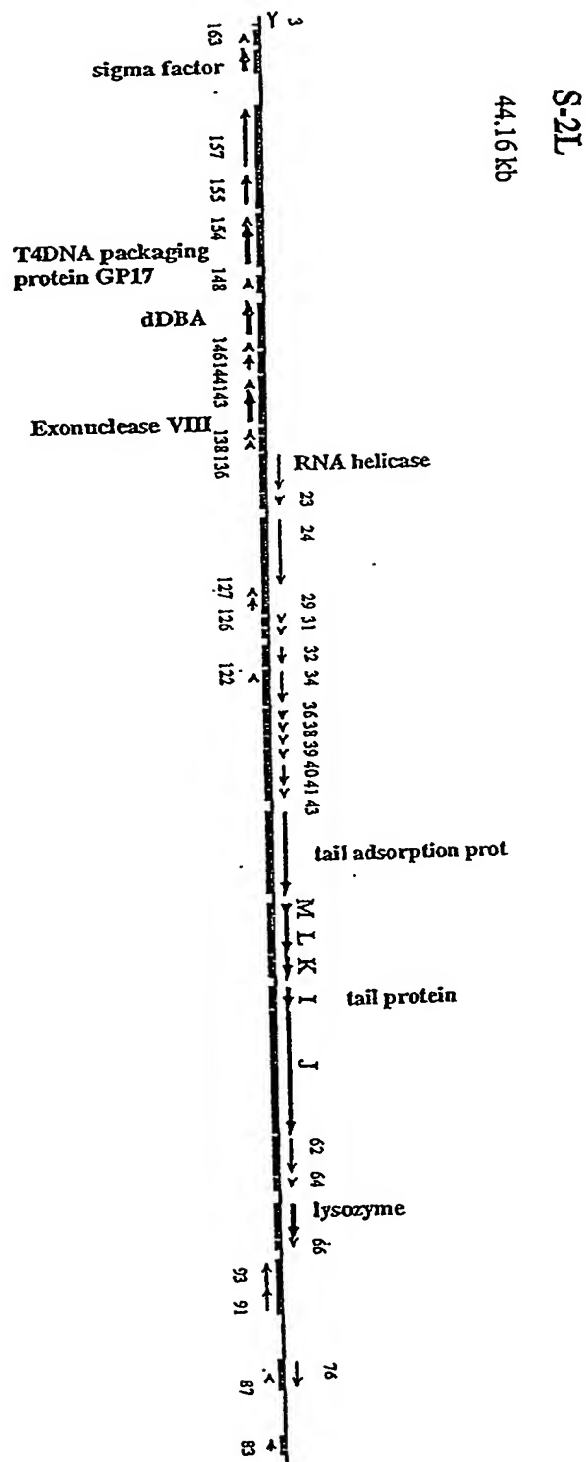
3/5

Figure 2b



4/5

Figure 3



5/5

Figure 4

